Identifying Response to Therapy in Longitudinal PET Imaging Studies

Phillips, Michael

Awarding institution:
King's College London

Identifying Response to Therapy in Longitudinal PET Imaging Studies

Michael Phillips

Division of Imaging Sciences & Biomedical Engineering

King's College London

A Thesis Submitted for the Degree of

Doctor of Philosophy

2014

Abstract

^{18}F-FDG PET can predict response using both qualitative and quantitative measures. PET Therapy Response Assessor (PETTRA) software was developed to allow users to view and analyse pre- and post- therapy images and compute quantitative measures for predicting response to therapy. Additionally, registration methodology was developed to register pre- and post- therapy PET/CT images. The methodology registers pre- and post- therapy PET/CT scans by registering CT scans using customised rigid and non-rigid registration performed by the Image Registration Toolkit (IRTK). Registration success was assessed using qualitative visual analysis and quantitative landmark analysis on a cohort of 20 lymphoma patients. Landmark analysis results found average misalignment on IRTK of ~10mm for rigid registration and ~6.5mm for non-rigid registration, in comparison with ~40mm with no registration applied. The effect of both rigid and non-rigid registration on transformed images was assessed. While rigid registration transformation caused minimal changes on intensity and tumour volume (<2%), non-rigid transformations caused changes of 11% and 21% respectively. PETTRA software was used to analyse quantitative parameters in 14 patients with mesothelioma and 85 patients with diffuse large B-cell lymphoma (DLBCL). For the 14 patients with mesothelioma, a range of parameters were used to assess response including SUV_{max}, SUV_{peak}, tumour volume (TV), total lesion glycolysis (TLG) and intensity volume histogram (IVH) parameters. TV and TLG were obtained using 13 fixed and 9 adapative threshold segmentation methods. Pre-and post- therapy SUV_{max}, SUV_{peak}, TV and TLG all showed promise in predicting survival. The comparison between TV and TLG obtained using different segmentation methods was negligible. For the 85 patients with DLBCL, SUV_{max}, TV and TLG struggled to predict response in patients according to ROC curves.

Acknowledgements

Firstly, I would like to thank my supervisors Prof. Paul Marsden and Prof. Derek Hill for their general support, ideas, feedback and time throughout the project. I am extremely grateful for the time and effort Dr. Sally Barrington put in to helping me segment hundreds of PET images and help refine the software made. Throughout the last month, I would like to thank Dr. Kiran Chana for her proof reading and general support. I would also like to thank Dr. David Cash, James Mackintosh, Dr. Edward Somer, Dr. Jane Mackewn, Dr. Paul Schleyer, Dr. Joel Dunn, Prof. Michael O'Doherty, Dr. Istvan Szanda, Dr. Tanuj Puri, Dr. Arunabha Roy, Dr. Anastasia Chalkidou and all the other staff who I have worked with in Imaging Sciences and at the St Thomas's PET Centre. I appreciate all the help I have had and your contributions have been invaluable to this work. I would also like to thanks all my friends and family who have supported me throughout what has been an enjoyable but demanding few years. I appreciate your patience and support. So many people have helped me throughout my time at KCL that I am sure I have forgotten some whose help has been priceless, I thank them also.

Publications

Conference Abstracts:

M. Phillips, E.J. Somer, P.J. Schleyer, J. Mackintosh, D.M. Cash, S.F. Barrington, M.J. O'Doherty, D.L.G. Hill, P.K. Marsden. Comparison of three registration methods for pre- and post- treatment oncological PET/CT scans. *Nuclear Medicine Communications*. 2010. 31 (5): 453-454.

M. Phillips, E.J. Somer, P.J. Schleyer, J. Mackintosh, D.M. Cash, S.F. Barrington, M.J. O'Doherty, D.L.G. Hill, P.K. Marsden. Evaluation of 3 registration methods for pre- and post-therapy PET/CT studies. *Journal of Nuclear Medicine*. 2010. 51 (Supplement 2): 1169.

M. Phillips, D.M. Cash, D.L.G. Hill, P.K. Marsden. Methodology for Identifying Response to Therapy in pre- and post- therapy PET/CT studies using Registered Images. *EANM Conference Abstract*. 2011.

Contents

List of Figures

List of Tables

Glossary of Abbreviations

Δ SUV_{max}	Change in SUV_{max} between pre- and post- therapy scans
2-D	2-dimensional
3-D	3-dimensional
aa-IPI	Age-adjusted International Prognostic Index
AC	Attenuation Correction
APD	Avalanche Photodiode
AIF	Arterial Input Function
AUC	Area Under the Curve
ASCT	Autologous Stem Cell Transplantation
BG	Background
BGO	Bismuth Germanate
BMI	Body Mass Index
BSA	Body Surface Area
^{11}C-MET	Methyl-[^{11}C]L-methionine
^{64}Cu-ATSM	Cu-diacetyl-bis(N^4-methylthiosemicarbazone)
CAD	Coronary Artery Disease
CC	Cross Correlation
CHOP	Cyclophosphamide, Doxorubicin, Vincristine and Prednisolone
CMR	Complete Metabolic Response
COM	Centre of Mass
CR	Complete Response
CRu	Complete Response (Unconfirmed)
CT	Computed Tomography
CTAC	Computed Tomography Attenuation Correction

CTV	Clinical Target Volume
DAR	Dose Absorption Rate
DEXA	Dual-Energy X-ray Absorptiometry
DICOM	Digital Imaging and Communications in Medicine
DLBCL	Diffuse Large B-Cell Lymphoma
DUR	Differential Uptake Ratio
EANM	European Association of Nuclear Medicine
ECOG	European Cooperative Oncology Group
EFS	Event-Free Survival
EORTC	European Organisation for Research and Treatment of Cancer
^{18}F-DOPA	^{18}F-fluorodihydroxyphenylalanine
^{18}F-FAZA	[^{18}F]fluoroazomycinarabinofuranoside
^{18}F-FCH	^{18}F-fluoromethylcholine
^{18}F-FDG	Flourine-18-labelled fluoro-2-deoxy-D-glucose
^{18}F-FES	16-α-[^{18}F]-fluoroestradiol
^{18}F-FET	O-(2-[^{18}F]fluoroethyl)-L-tyrosine
^{18}F-FLT	3-deoxy-3-^{18}F-flurothymidine
^{18}F-FMAU	2'-[^{18}F]fluoro-2'-deoxy-5-methyl-1-β-arabinofuranosyluracil
^{18}F-MISO	[^{18}F]fluoromisonidazole
FBP	Filtered Back Projection
FDG	fluoro-2-deoxy-D-glucose
FLAB	Fuzzy Locally Adaptive Bayesian
FLE	Fiducial Localisation Error
FWHM	Full Width Half Maximum
^{68}Ga-DOTATOC	^{68}Ga-labeled [1,4,7,10-tetraazacyclododecane-N,N′,N″,N‴-tetraacetic-acid-DPhe1-Tyr3]-octreotide

GE	General Electric
GIPL	Guy's Image Processing Lab
GLCM	Grey Level Co-occurrence Matrix
GLUT	Glucose Transporters
GSO	Germanium Oxyorthosilicate
GTM	Geometric Transform Matrix
GTV	Gross Tumour Volume
GUI	Graphical User Interface
GUIDE	Graphical User Interface Design Component
HL	Hodgkin's Lymphoma
IDIF	Image Derived Input Function
IMRT	Intensity Modulated Radiotherapy
IPI	International Prognostic Index
IRTK	Image Registration Toolkit
IVH	Intensity Volume Histogram
IWC	International Working Group Criteria
$LaBr_3$	Lanthanum Bromide
LBM	Lean Body Mass
LGI	Larson-Ginsberg Index
LOR	Line of Response
LSO	Lutetium Oxyorthosilicate
LYSO	Lutetium-yttrium Oxyorthosilicate
MB	Metabolic Burden
MCT	Multi-Centre Trials
MLEM	Maximum Likelihood Expectation Maximisation
MNL	Maximum Normal Level

MPM	Malignant Pleural Mesothelioma
MR	Magnetic Resonance
MR_{glu}	Glucose Metabolic Rate
MRI	Magnetic Resonance Imaging
MTB	Metabolic Tumour Burden
MTV	Metabolic Tumour Volume
MWPC	Multiwire Proportional Chambers
NaI (Tl)	Sodium Iodide doped in Thallium
NCI	National Cancer Institute
NE	Non-Evaluable
NECR	Noise Equivalent Count Rate
NEMA	National Electrical Manufacturers Association
NHL	Non-Hodgkin's Lymphoma
NLR	Nonlinear Regression
NMI	Normalised Mutual Information
NSCLC	Non-Small Cell Lung Cancer
OS	Overall Survival
OSEM	Ordered Subsets Expectation Maximisation
pcc	Pearson Correlation Coefficient
PD	Progressive Disease
PERCIST	PET Response Criteria for Solid Tumours
PET	Positron Emission Tomography
PETTRA	Positron Emission Tomography Therapy Response Assessor
PFS	Progression Free Survival
PMD	Progressive Metabolic Disease
PMR	Partial Metabolic Response

PMT	Photomultiplier Tube
PR	Partial Response
PSF	Point Spread Function
PV	Partial Volume
QC	Quality Control
R-CHOP	Rituximab with CHOP
R-CEOP	Rituximab with Cyclophosphamide, Etoposide, Prednisolone, Vincristine
RC	Recovery Coefficient
REAL	Revised European-American Classification of Lymphoid Neoplasms
RECIST	Response Evaluation Criteria in Solid Tumours
R-IPI	R-CHOP version of International Prognostic Index
ROC	Receiver Operator Characteristics
ROI	Region of Interest
RTL	Relative Threshold Level
S/B	Source to Background
S.D.	Standard Deviation
SD	Stable Disease
SKM	Simplified Kinetic Method
SMD	Stable Metabolic Disease
SNR	Signal to Noise Ratio
SPECT	Single Photon Emission Computed Tomography
SSD	Sum of Squared Differences
SUL	Standardised Uptake Value normalised by Lean Body Mass
SUV	Standardised Uptake Value
SUV_{BSA}	Standardised Uptake Value normalised by Body Surface Area

$SUV_{LBM/lean}$	Standardised Uptake Value normalised by Lean Body Mass
SUV_{m}	Standardised Uptake Value normalised by a weight index
SUV_{max}	Maximum Standardised Uptake Value
SUV_{mean}	Mean Standardised Uptake Value
SUV_{peak}	Peak Standardised Uptake Values
SV40	Simian Virus
SWOG	South West Oncology Group
$t_{½}$	Half Life
T	Tumour
T/B	Tumour to Background
TGA	Total Glycolytic Activity
TGV	Total Glycolytic Volume
TLG	Total Lesion Glycolysis
ToF	Time of Flight
TRE	Target Registration Error
TTP	Time to Progression
TV	Tumour Volume
VEGF	Vascular Endothelial Growth Factor
VOI	Volume of Interest
WHO	World Health Organisation

1) Introduction and Background

1.1) Positron Emission Tomography

1.1.1) Overview of PET

Positron emitters were first used in medical research as early as the 1950s to try and localise brain tumours (Brownell and Sweet, 1953), with the first Positron Emission Tomography (PET) scanners developed in the 1970s (Ter-Pogossian *et al*., 1975; Phelps *et al*., 1975). Advances in technology and development of other tracers, most notably fluoro-2-deoxy-D-glucose (FDG) (Reivich *et al*., 1979), have gradually increased research in PET and its use clinically. In the last ten years, the introduction of scanners combining PET with computed tomography (CT) have resulted in a surge of interest in the modality (Beyer *et al*., 2000), primarily for cancer imaging, due to the ability of PET to show functional *in vivo* processes compared with anatomical information (Juweid and Cheson, 2006). PET has low spatial resolution (~5mm) in comparison to other imaging modalities so the complementary combination of metabolic PET images and higher resolution anatomical CT images, which typically have a spatial resolution of <1mm, make it a very powerful and useful imaging tool.

PET works by injecting a radionuclide tracer into a patient emitting positrons within the body. These positrons travel a short distance known as the positron range, dependent upon the isotope but usually <1mm for ^{18}Flourine, before undergoing annihilation with an electron producing two 511keV gamma (γ) rays which can be absorbed by detectors (Figure 1.1). If two γ rays are detected at 180° within a certain time they are deemed to represent a 'coincidence' or 'true' event. These can be tracked to a position somewhere along the line of response (LOR) joining them. Detectors absorb emitted γ rays from these events and produce a signal containing enough

information to determine when the event occurred and how much energy it deposited. Using the information from many events, an image of tracer uptake can be reconstructed.

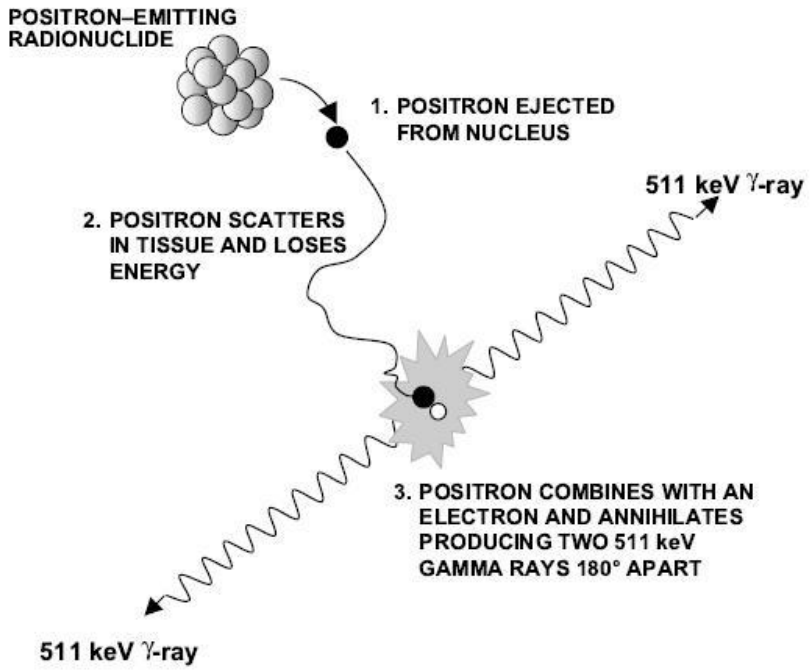


Figure 1.1: Positron Annihilation

A positron is emitted from the nucleus of a radionuclide and annihilates with an electron producing two 511 keV γ rays 180° apart (Taken from Cherry, 2001).

1.1.2) PET Noise and Spatial Resolution

Not all detected coincidence events are from the same 'true' annihilation as scatter and random coincidences can be detected causing noise in an image. Scatter coincidences occur when γ rays undergo Compton scattering and end up in different positions to what would be expected, resulting in a false LOR. Random coincidences occur when two coincidences are detected within the same time frame and are therefore deemed to represent a true event, when in fact they are not from the same annihilation (Figure 1.2).

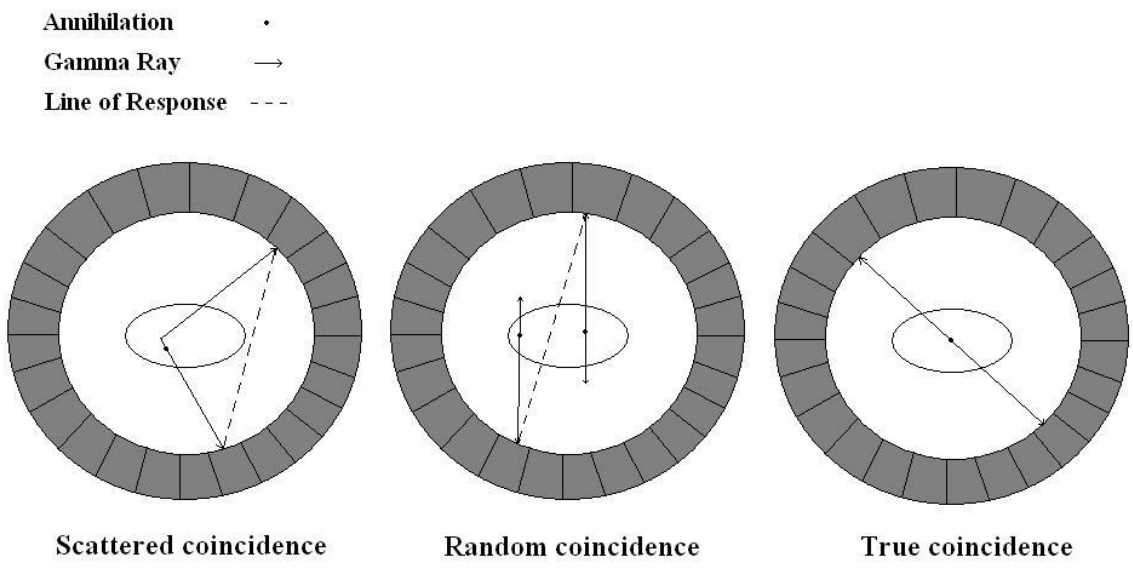


Figure 1.2: Scattered, Random and True Coincidence Events

Scattered and random coincidences cause noise in an image as true coincidences are ideally the only event that PET should detect (Adapted from Rohren *et al*., 2008).

To gauge the effect of random and scattered coincidences, Strother *et al.* (1990) introduced the Noise Equivalent Count Rate (NECR), designed to measure performance of an imaging system in terms of a signal-to-noise ratio (SNR), i.e. the amount of desired signal compared to unwanted noise (Strother *et al.*, 1990). NECR is mathematically defined as:

$$NECR = \frac{\text{True Events}}{\text{True Events} + \text{Random Events} + \text{Scatter Events}} \quad [1.1]$$

Physical factors can affect spatial resolution by up to 2mm, but clinical scanners have a resolution of ~6-8mm (Townsend, 2004; Papathanassiou *et al*., 2009). Spatial resolution is affected by positron range, acollinearity of photons and decoding of the PET signal. The positron range results in the LOR being recorded at the point where the positron was annihilated, not from where it was emitted. The assumption that annihilation photons are ejected at exactly 180° is only approximate as they are slightly acollinear, at an average angle of 0.25°, causing a

difference in the annihilation point to the LOR. Photon non-collinearity occurs because a positron's centre of mass is not always at rest at the point of annihilation and in order to conserve energy and momentum, the annihilation photons are not directly 180° apart. Additionally, in decoding the PET signal, there are sampling errors and penetration of γ rays on more than one detector element. While these issues adversely affect the spatial resolution, the most dominant factor is the size of detector elements (Moses, 2011).

1.1.3) PET Detectors

There are two main types of PET detectors. The first uses separate elements, consisting of a scintillation crystal and a photomultiplier tube (PMT), to deduce the position of the photon from where it hits the detector. The second comes in the form of a gamma camera-like system so the position of the photon can be obtained from signals processed by PMTs (Rohren *et al*., 2004). Most installed systems are block detectors (Casey and Nutt, 1986), a mixture of the two designs, which are arranged around the patient in a ring configuration and contain blocks of scintillator segmented into an array read out by PMTs (Figure 1.3).

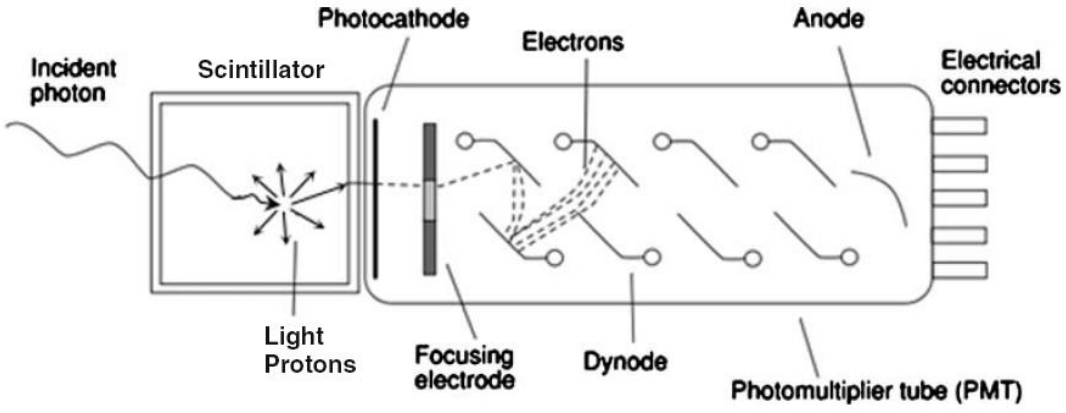


Figure 1.3: Diagram of a Standard PET Detector

Photons are absorbed by scintillators, which are attached to PMTs converting scintillation photons into electrical currents (Taken from Lecomte, 2009).

Other detector designs have been developed in an attempt to improve PET, including PETRRA positron cameras which use large area multiwire proportional chambers (MWPC) coupled with scintillators (Duxbury *et al*., 1999), detectors using avalanche photodiode (APD) devices with higher gain (~10^2-10^3) and faster timing (1ns) (Herbert *et al*., 2007), and silicon detectors using metallic strips and amplifiers (Park *et al*., 2007). Five key attributes an ideal detector should have are: (i) high stopping power i.e. ability to absorb γ rays, (ii) high spatial resolution to determine the exact location of the γ ray in the spatial volume, (iii) good energy resolution to stop scattered events, (iv) high temporal resolution and (v) low cost (Lewellen, 2008). Early PET detectors used thallium doped sodium iodide (NaI(Tl)) as scintillation crystals, however, bismuth germanate (BGO) has been most commonly used in the last two decades due to its high density and stopping power (Lecomte, 2009). However, lutetium oxyorthosilicate (LSO) and germanium oxyorthosilicate (GSO) are now being used in PET systems and are beginning to replace BGO (Patel *et al*., 2010).

To achieve a spatial resolution of ~2mm, detector elements need to be <4mm in thickness so development of smaller elements would result in a higher spatial resolution (Cherry, 2006). Detectors with extremely good temporal resolutions, in the range of 500-900 picoseconds (Spanoudaki and Levin, 2010), can allow time of flight (ToF) PET imaging, which calculates the time difference between detected annihilation photons to approximate where annihilation occurs along the LOR. ToF PET's approximate improvement in SNR is:

$$SNR_{TOF} \approx \sqrt{(2D / c\Delta t) * SNR_{non\text{-}TOF}} \qquad [1.2]$$

where D is the diameter of the imaged object, c is the speed of light and Δt is the timing resolution of the system (Cherry, 2006). ToF imaging has shown to significantly improve tumour detection (Kadrmas *et al*., 2009; Surti *et al*., 2011; El Fakhri *et al*., 2011). The main

reason for increased availability and effectiveness of ToF imaging is the development of scintillators with high timing resolutions such as LSO, lutetium-yttrium oxyorthosilicate (LYSO) and lanthanum bromide ($LaBr_3$) crystals (Peng and Levin, 2010).

1.1.4) PET Reconstruction

PET reconstruction takes raw imaging data and produces a 3-D image. Raw data can be formatted into a sinogram, formed by projections of LORs from different angles, and an image can be created using a reconstruction algorithm. The first reconstruction method to gain popularity was filtered back projection (FBP), which filters sinogram data before using backprojection to create an image. Iterative methods of reconstruction have now replaced analytical methods, such as FBP, demonstrating improved image quality and better tumour detectability (Meikle *et al*., 1994; Hutton, 2011). Iterative algorithms start with an initial guess of distribution which is then projected according to scanner geometry and compared to measured projections. The difference between the initial guess and measured projections is used to continue the iterative process until they agree within their statistics (Ziegler, 2005). One advantage of using iterative algorithms is that *a priori* information, such as noise, attenuation and detector nonuniformity, can be incorporated for more accurate reconstruction (Tarantola *et al*., 2003). The most popular method of reconstruction is the maximum likelihood expectation maximisation algorithm (MLEM) (Shepp and Vardi, 1982), and its accelerated form, the ordered subsets expectation maximisation (OSEM) method which breaks down projection data into subsets to reduce computational burden (Hudson and Larkin, 1994).

1.1.5) PET Attenuation Correction

Attenuation occurs when photons are absorbed by tissue before reaching a detector causing structures deep in the body to have falsely low tracer uptake. To eradicate this problem, attenuation correction (AC) can be performed by applying a correction factor to the number of

events recorded within each LOR. For example, if five events are counted between two detectors and it is deduced that only 10% of photon pairs survive, then it can be corrected so 50 counts are recorded rather than five (Rohren *et al*., 2004). This can be achieved by taking a 'transmission' scan where the scan is acquired using an external radionuclide point source with the patient in scanning position. CT attenuation correction (CTAC) is used on PET/CT scans as it's faster, contains more data and produces less statistical noise (Kinahan *et al*., 1998). This comes at a cost as CTAC can compromise the accuracy of PET data as respiratory motion can cause misalignment between CTAC and PET scans as they are acquired at different times and speeds. This can result in artefacts around the dome of the liver and the lung on AC PET images in some cases (Bacharach, 2006). There has been significant work to overcome this issue including averaging CT data over respiratory cycles (Pan *et al*., 2005; Chi *et al*., 2007), and respiratory gating CT and PET data (Nehmeh *et al*, 2004; McQuaid *et al*., 2011).

1.1.6) PET Tracers for Oncology

PET tracers are labelled with positron-emitting radionuclides which indirectly cause production of back-to-back γ rays. The most common radionuclides in PET imaging studies are oxygen-15, nitrogen-13, carbon-11 and fluorine-18. They have half-life's ($t_{½}$) of 2min, 10min, 20min and 110min respectively making them suitable for imaging. Their $t_{½}$'s are long enough to allow imaging to take place while being short enough to not give unnecessary doses of radiation to a patient. Radiotracers/radiopharmaceuticals for PET imaging are synthesised from positron emitting radionuclides which are usually produced in a cyclotron where stable elements are bombarded with protons or negative ions (Schleyer, 2004). Fluorine-18 is commonly the radionuclide of choice as it has an ideal $t_{½}$ to allow time for imaging, produces a high positron yield (~97%) and low positron energy. However, it does have complicated radiochemistry and is not a complete, isotropic tracer as there is no natural fluorine in organic compounds. Therefore, Carbon-11 is of particular interest since carbon is heavily involved in the chemical structure of

organic compounds, however its shorter $t_{½}$ of 20min limits the time available for synthesis, and its integration requires skill and experience (Wiebe, 2004).

Flourine-18-FDG (^{18}F-FDG) is by far the most widely used tracer in PET imaging measuring glucose metabolism prevalent in fast growing cancer cells, however, undesirably, there is also normal physiological uptake in organs including the brain, kidneys and bladder (Warburg, 1956). ^{18}F-FDG also accumulates in areas of inflammation and infection and it shows relatively low uptake in some diseases such as prostate cancer, neuroendocrine tumours and hepataocellular carcinomas that may replace or compliment FDG (Mercer, 2007). This highlights the need for extensive ongoing research into other tracers. One of the most promising groups of tracers are markers for cell proliferation i.e. cell growth and division. Nucleoside analog 3-deoxy-3-^{18}F-flurothymidine (^{18}F-FLT) shows good correlation with tumour proliferation making it an ideal evaluation tool for therapeutic response (Shields *et al*., 1996; Shields *et al*., 1998; Reske and Deisenhofer, 2006). Another nucleoside analog is 2'-[^{18}F]fluoro-2'-deoxy-5-methyl-1-β-arabinofuranosyluracil (^{18}F-FMAU) while amino acids such as Methyl-[^{11}C]L-methionine (^{11}C-MET) and O-(2-[^{18}F]fluoroethyl)-L-tyrosine (^{18}F-FET) also show uptake representing cell proliferation (Dunphy and Lewis, 2009). Choline kinase activity and phospholipid synthesis targeted tracers can also show proliferating tissues and tracers like ^{11}C-choline, ^{11}C-acetate and ^{18}F-choline analogs, such as ^{18}F-fluoromethylcholine (^{18}F-FCH), are of potential future use (DeGrado *et al*., 2001; Pantaleo *et al*., 2008).

Hypoxia tracers are another important area of research. Hypoxia has been found to increase the likelihood of tumour progression and failure of radiotherapy due to increased radioresistance of hypoxic cells compared with oxygenated ones (Vaupel and Mayer, 2007; Overgaard, 2007; Lucignani, 2008). Three hypoxia tracers showing particularly promising uptake in cells with hypoxia are [^{18}F]fluoromisonidazole (^{18}F-MISO) (Rasey *et al*., 1996),

[^{18}F]fluoroazomycinarabinofuranoside (^{18}F-FAZA) (Sorger *et al*., 2003), and Cu-diacetyl-bis(N^4-methylthiosemicarbazone) (^{64}Cu-ATSM) (Lewis *et al*., 1999). Examples of other tracers being researched in PET imaging are ^{18}F-annexin V for targeting apoptotic changes in cell death, ^{18}F-fluorodihydroxyphenylalanine (^{18}F-DOPA) and ^{11}C-5-hydroxytryptophan for targeting amine transport common in peptide producing neuroendocrine tumours, and cellular receptors 16-α-[^{18}F]-fluoroestradiol (^{18}F-FES) as an estrogen receptor for breast cancer imaging and ^{68}Ga-labeled [1,4,7,10-tetraazacyclododecane-N,N′,N″,N‴-tetraacetic-acid-DPhe1-Tyr3]-octreotide (^{68}Ga-DOTATOC) as a somatostatin receptor to image neuroendocrine tumours (Mercer, 2007).

1.1.7) Clinical Applications of PET

PET has a wide variety of applications in medical imaging. It was first used in neurology as early as the 1980s to investigate movement disorders such as Huntingdon's disease and Parkinson's disease (Kuhl *et al*., 1982; Raichle *et al*., 1984). Since then, it has also been used to study dementia (Alzheimer's disease), brain tumours (gliomas) and epilepsy with a variety of tracers (Tai and Piccini, 2004). PET is used in cardiology to detect coronary artery disease (CAD), due to its higher spatial resolution and heart-to-background ratio when compared to single photon emission computed tomography (SPECT) (Knuuti, 2008). PET can also assess myocardial perfusion, blood flow, heart failure and identify plaque burden (Maisey, 2002; Schwaiger *et al*., 2005).

PET is most widely used in oncology. Up to 90% of PET investigations are performed using ^{18}F-FDG for cancer imaging (Lonsdale and Beyer, 2010). Its huge growth in this area is due to its importance in cancer management in which it can be used to confirm suspected malignancies, determine staging and site of disease, diagnose recurring disease and assess response to therapy (Maisey, 2002). ^{18}F-FDG PET is widely used clinically and has large practical usefulness in diagnosing head and neck and lung cancers, as well as staging lung, oesophageal, colorectal,

lymphoma, melanoma, and head and neck cancers. ^{18}F-FDG PET also has the ability to detect recurrence in melanoma, colorectal, breast, ovarian and head and neck cancers. It is being increasingly used in these areas for other cancers such as pancreatic, uterine, testicular, thyroid, sarcoma and gastrointestinal stromal (Fletcher *et al*., 2008; Papathanassiou *et al*., 2009). A growing area of clinical PET/CT is identifying response to therapy in patients with a variety of different cancers, particularly lymphomas, non-small cell lung cancer (NSCLC), breast, colorectal and oesophageal cancers (Ben Haim and Ell, 2009).

^{18}F-FDG PET is also useful for the investigation of primary lesions of unknown cancer, which occurs in 2-4% of all cancer patients (Jerusalem *et al*., 2003). While ^{18}F-FDG PET is very useful for imaging many cancers including gastric carcinomas, bronchialveolar cell carcinomas, mucosa-associated lymphoid tissue lymphomas, small lymphocytic cell lymphoma, carcinoids and renal cell carcinomas do not take enough ^{18}F-FDG to be used accurately (Lucignani *et al*., 2004). ^{18}F-FDG PET has been shown to have a positive and significant impact on therapeutic management of cancer patients in oncology (Zafra *et al*., 2008).

PET has been highlighted as a useful tool in drug development and pharmacology as it can detect early signs of drug success or failure and could provide the most superior and early assessment of drug efficacy in comparison to current methods (Kelloff *et al*., 2005). The main benefit of PET is that it can show, very early, the effectiveness of a given drug while also being useful in pre-clinical imaging for better translation of animal models to clinical human patients (Pien *et al*., 2005). ^{18}F-FDG PET has been used to assess traditional chemotherapy agents in breast cancer, colorectal cancer and melanoma, as well as novel agents such as epidermal growth factor receptors, successfully detecting response within the first few weeks of treatment (Hammond *et al*., 2003). It can also measure the uptake of a drug to evaluate its bioavailability

i.e. whether the tumour takes up enough concentration of a drug to kill cancer cells whilst most normal tissue remains unharmed.

Recent research has shown increased use of PET in areas including paediatric oncology and radiotherapy planning where images can be registered and segmented to establish a gross tumour volume (GTV) to be irradiated (Nestle *et al*., 2009; Bussink *et al*., 2010; London *et al*., 2010; Portwine *et al*. 2010, Lucignani and De Palma, 2011; De Ruysscher *et al*., 2012). Another clinical application of PET is its emerging use in combination with magnetic resonance imaging (MRI), an imaging modality which images the nuclei of atoms using magnetic fields and radiowaves. PET/MRI offers benefits over PET/CT including more reliable image registration and lower radiation exposure, although there are limitations in its development including issues with AC and in what areas it may be of best use (Schiepers and Dahlbom, 2011).

1.1.8) PET Imaging Protocol and Quality Control

The standard protocol for clinical whole body imaging with ^{18}F-FDG PET starts with an injection of ~370MBq of ^{18}F-FDG. The patient is then left to rest for a period of ~1h to allow for tracer uptake. Images are acquired in 5-9 bed positions of ~15cm, taken sequentially for 2-5min per section (Marsden and Sutcliffe-Goulden, 2000). Imaging protocols can differ depending on factors including breath holding, contrast agents, CT operating parameters, PET scan time, and optimal injected dose. In monitoring response to therapy in clinical studies, care must be taken to make sure these factors are as similar as possible in each scan to enhance repeatability.

A documented quality control (QC) assurance program should be carried out at all institutions for PET scanners including annual/monthly testing of hardware/software and daily maintenance checks for detector performance (using a transmission scan) and image uniformity (Zanzonico, 2008). There are a number of defined measurement standards which can be followed for

performance measures for spatial resolution, sensitivity, scatter fraction, count losses and image quality e.g. National Electrical Manufacturers Association (NEMA) NU 2-2001 PET performance measures (Daube-Witherspoon *et al*., 2002). For clinical trials it is normally expected that a full QC program is completed and properly documented, particularly for multi-centre trials (MCTs) where different scanning procedures are likely to increase variation in SUVs (Takahashi *et al*., 2008). At St Thomas' Hospital, QC procedures are based on those of the American College of Radiology Imaging Network. This requires making sure the specified technical scanner specifications and acquisition and reconstruction protocols are used, and nominating people responsible for scanning and QC at each institution to ensure that QC tests are conducted. In addition, correct data transfer methods and anonymisation of patient data are required. The PET Imaging Centre at St Thomas' Hospital is the main base for the National Cancer Institute Research clinical trials network in the UK.

1.2) Response Assessment in PET

1.2.1) Introduction to Response Assessment

^{18}F-FDG PET was first studied in assessing response in the early 1990s in a group of breast cancer patients undergoing chemohormonotherapy (Wahl *et al*., 1993). Further studies have shown that PET can identify whether treatment is working after days to weeks, rather than weeks to months compared with CT imaging (Wieder *et al*., 2005). This has resulted in PET/CT being increasingly used to investigate response to therapy in a number of different cancers, particularly lymphoma (Brepoels *et al*., 2007; Ben Haim and Ell, 2009). Comparing a pre-therapy scan with a post-therapy scan within a few weeks of treatment can accurately predict response to chemotherapy (Kostakoglu *et al*., 2002). Identifying whether tumours will respond within the

first few weeks of treatment is of great clinical importance as it means non-responding patients do not have to undergo unsuccessful therapy reducing side effects and cost (Weber, 2005).

Many criteria have been devised to assess response, mainly for anatomical measurements on CT, but criteria for metabolic changes on PET are now being included due to the disadvantages of anatomical imaging modalities and growing evidence that ^{18}F-FDG uptake is a predictor of tumour response (Larson and Schwartz, 2006). Qualitative visual interpretation is used in the clinical environment along with semi-quantitative measures including standardised uptake values (SUVs) to quantify the amount of ^{18}F-FDG uptake in a lesion. Semi-automated techniques for identifying response are of interest, as is research into analysing scans. A robust, accurate and easy to use method could potentially be a huge benefit in clinical studies and for diagnostic use.

1.2.2) Qualitative Methods

1.2.2.1) Visual Assessment

Visual interpretation of ^{18}F-FDG PET images is used clinically and has been shown to adequately identify patients who are likely to respond to current therapies (Miller *et al*., 2003). A study assessing 161 pulmonary nodules showed that visual assessment was as accurate as semi-quantitative methods at distinguishing between positive and negative lesions with only faintly positive lesions showing better evaluation when using quantifiable SUV values (Nomori *et al*., 2005). Despite its simplicity, observer variability in the reporting of PET scans is seen as an issue as it is not quantifiable. Semi-quantitative methods, such as SUVs, are considered more robust and reliable measures at predicting response in lymphoma (Lin *et al*., 2007), however, visual analysis is still used clinically and has the advantages of more intelligent interpretation of images. While SUVs may quantify changes in uptake, an experienced oncology clinician will

always be more valuable when assessing response as they can recognise where FDG uptake is likely to be physiological, if new lesions have appeared, and whether changes in uptake in some anatomy are more meaningful than in other areas.

1.2.2.2) Anatomical Guidelines

In 1981, the World Health Organisation (WHO) proposed a set of guidelines, primarily designed for use with X-ray and CT, suggesting the use of bidimensional measurements for identifying tumour response. Bidimensional measurements are obtained by taking the longest diameter of each tumour in one direction and multiplying this with the longest perpendicular diameter, the product of these measurements for each tumour can then be used to compare pre- and post-therapy results (Miller *et al*., 1981). Other groups, such as the South West Oncology Group (SWOG) in the United States (Green and Weiss, 1992), concurred with the WHO guidelines in terms of tumour measurements but suggested subtle differences in terms of response criteria, such as using different changes in volume to determine response categorisation. WHOs bidimensional method of measurement has been criticised, as the process of measuring in two dimensions and calculating the products is arduous and still has potential for error (Oh Park *et al*., 2003). New guidelines were introduced by the Response Evaluation Criteria in Solid Tumours (RECIST) Group in the late 1990s which attempted to update and simplify the WHO methods, most notably by proposing unidimensional measurements of tumours, taking the maximum 2-dimensional (2-D) diameter of each tumour in a 3-dimensional (3-D) image (Therasse *et al*., 2000).

Studies have shown that RECIST correlates well with the WHO criteria, validating the more simplistic method of measurement as a time saving measure (James *et al*., 1999; Trillet-Lenoir *et al*., 2002). Other changes in the criteria have seen the percentage changes needed in tumour

volume (TV) for different categories of response modified, meaning that there can be discordance between different criteria (Julka *et al*., 2008) (Table 1.1). There have also been criteria developed for specific cancers, for example, the International Working Group criteria (IWC) for non-Hodgkin's lymphoma (NHL) (Cheson *et al*., 1999).

Response	WHO	SWOG	RECIST
Complete Response (CR)	Disappearance of all known disease	Complete disappearance of all measurable disease	Complete disappearance of all target and non-target lesions
Partial Response (PR)	Decrease of >50% in TV, no new lesions found	Decrease of >50% in TV, no new lesions found	Decrease of >30% in TV, no new lesions found
Stable Disease (SD)	Decrease of <50% or increase of <25% in TV	Decrease of <50% or increase of <50% in TV	Decrease of <30% or increase of <20% in TV
Progressive Disease (PD)	Increase of >25% on smallest measurement of TV or new lesion(s) found	Increase of >50% or >10cm^2 in TV or new lesion(s) found	Increase of >20% over smallest sum of the max diameter or new lesion(s) found

Table 1.1: Comparison of Anatomical Response Criteria

WHO, SWOG and RECIST criteria have the same four categories for CR, PR, SD and PD with slight differences in what % changes in tumour volume (TV) warrants each type of response.

The most recent adaptation of anatomical response criteria, RECIST 1.1 (Eisenhauer *et al*., 2009), proposes reducing the number of lesions used for calculating total TV from a maximum of 10 to 5, improving simplicity without significantly affecting response categorisation. While all aforementioned anatomical response criteria may have developed better, more substantial and easier to follow guidelines over time, they all suffer from the same flaws when it comes to assessing response. Firstly, they rely heavily on observers measuring lesions accurately, a task which can result in significant inter- and intra- observer variability potentially causing inconsistency and incorrect interpretation of tumour response (Thiesse *et al*., 1997; Erasmus *et al*., 2003). Secondly, they do not take into account metabolic processes including necrosis, cavitation and fibrosis which can occur in tumours often meaning that a reduction in size is slow

or non-existent (Hicks, 2005). Finally, they do not consider metabolic imaging techniques, such as PET, which could potentially solve these issues.

An increasing number of drugs in development do not kill cells like traditional cytotoxic treatments, but stop tumour cell growth (cytostatic treatment). This results in a constant tumour size, even if therapy is successful, meaning that anatomical TV measurements cannot be used accurately as an indicator of response (Tuma, 2006). PET can show changes in cell growth faster than anatomical imaging modalities and, therefore, is gradually being used more to identify response. A CT criteria based on both tumour size and density, based on Hounsfield numbers on contrast enhanced CT, has been suggested as a possible solution to account for changes in tumour density and provide similar results to PET response (Choi *et al*., 2007). However, it is clearly unlikely to compete with PET on a long term basis as PET has more functional imaging capabilities.

1.2.2.3) Metabolic Guidelines

The continuing use of PET/CT in clinical situations along with a wide range of literature proving its success at identifying response to therapy, in a variety of different cancers, has meant that anatomical guidelines need to be updated for PET/CT (Avril *et al*., 2009; Ben Haim and Ell, 2009; Hutchings and Barrington *et al*., 2009; Krause *et al*., 2009; Schöder *et al*., 2009; Schwarz *et al*., 2009). The increased use of PET has led to criteria being redesigned for PET/CT imaging, for example, revised IWC integrating ^{18}FDG-PET for response assessment in NHL (Juweid *et al*., 2005), and new criteria being developed for assessing and categorising response to therapy using PET. The first criteria for PET imaging was developed by the European Organisation for Research and Treatment of Cancer (EORTC) PET study group and made recommendations on patient preparation, scanning procedure and timings as well as response categories (Young *et al*.,

1999) (Table 1.2). Other studies have continued to make recommendations on standards for PET imaging (Shankar *et al*., 2006; Delbeke *et al*., 2006; Boellaard *et al*., 2010). However, the only other guidelines to specifically look at response are the PET Response Criteria for Solid Tumours (PERCIST) (Wahl *et al*., 2009). The criteria use the same four response categories as EORTC criteria but are slightly updated to represent developments over the decade since EORTC were developed (Table 1.2).

Response	EORTC	PERCIST
Complete Metabolic Response (CMR)	Complete resolution of FDG uptake within TV (Tumour Volume)	Complete resolution of FDG uptake within measurable target lesion, indistinguishable from background levels
Partial Metabolic Response (PMR)	A reduction >15% in tumour SUV	Reduction of a >30% in SUL_{peak}, absolute drop must be >0.8 units
Stable Metabolic Disease (SMD)	Increase in tumour SUV of <25% or decrease of <15% and no visible increase in uptake	No more than a 30% increase or decrease in SUL_{peak}. Not CMR, PMR or PMD
Progressive Metabolic Disease (PMD)	Increase in tumour SUV of >25%, viable increase of uptake (20% in longest dimension) or uptake in a new lesion(s) found	Increase of >30% in SUL_{peak}, absolute increase >0.8 units. Alternatively, a visible increase in overall uptake (>75% change in TV) or new lesion(s) found

Table 1.2: Comparison of Metabolic Response Criteria

EORTC criterion and PERCIST have the same four categories of response: CMR, PMR, SMD and PMD. However, like anatomical criteria, each has differences in how change is measured and what warrants which response. SUL = SUV normalised using lean body mass rather than body weight.

The PERCIST criterion has many similarities to RECIST as it uses similar data analysis methods, such as analysing up to five tumours, however, it does highlight the limitations of categories of response and claims that data is intrinsically continuous and should be treated as such. The main differences come in the guidelines for the PET scan itself and methods of quantification used in determining response to therapy which are different to the anatomical

measurements in RECIST. The maximum SUV in the tumour (SUV_{max}) is recommended as the main quantitative measure to assess response but SUV_{peak}, an average SUV of a small spherical volume around the maximum pixel/voxel, and metabolic tumour volume (MTV) measures combined with SUVs are suggested as possible measures for the future. The criteria also suggests SUV should be corrected for lean body mass (SUL) rather than body weight. The invention of PERCIST is a clear sign of the need for specific PET guidelines to analyse response to therapy. The standardisation of PET scanning to minimise variability of SUVs and produce reliable results relates heavily to this and is a major issue in quantification and response and is discussed in more detail in 1.2.6.

1.2.3) Semi-Quantitative Methods

Semi-quantitative methods measure the uptake of tracer in a PET image so values are relative to radionuclide concentration. Quantification of uptake can be calculated in an image in which pixel values are proportional to Bq/ml. This can be taken one step further by obtaining physiological quantification of specific biological processes such as the measurement of perfusion in $ml/min^{-1}/g^{-1}$, glucose metabolism in $mol/min^{-1}/g^{-1}$ and parameters relating to receptor properties such as density and occupancy (Marsden, 2004). It has been shown that for assessing FDG uptake in tumours, using semi-quantitative measures on static scans is adequate and avoids unnecessary blood sampling (Minn *et al*., 1993).

Over two decades ago, Di Chiro and Brooks (1988) argued that quantification was not as important as many deemed it and that visual inspection was more accurate (Di Chiro and Brooks, 1988). PET imaging and related technologies have progressed since and it has been suggested that although quantification is not a necessary tool, it can provide useful information for newer observers of images to make a decision on whether a lesion is benign or malignant (Coleman, 2002). It has been stated that a key role of quantitation is identifying response to

therapy as it can be a useful tool to the oncologist in deciding whether a tumour is responding to treatment (Graham, 2002). Accurate quantitative analysis needs PET images which have had emission data corrected for random and scatter coincidences, attenuation, differences in detector efficiencies and detector geometry and dead-time effects (Visvikis *et al*, 2004).

1.2.3.1) Tumour to Background Ratio

The tumour to background ratio (T/B) ratio is an index which can be calculated from a static PET scan. It is a comparative guide which uses normal tissue, such as the liver or mediastinum, as a background measure to distinguish areas of high activity and possible tumours. However, as a result of using background values, it is subject to being affected by changes in normal tissues and is therefore regarded as an unreliable index compared to other semi-quantitative methods (Castell and Cook, 2008). It is rarely seen as an index for assessing response in current research, however, one study using contrast ratios of the brain and lung compared to tumours, found them to provide better results compared to SUV_{max} (Nomori *et al*., 2005). Highest activities in the tumour (T) and background (B), using either the lung or brain, were measured and the contrast ratio was calculated as (T-B) / (T+B) in each nodule as an index for uptake.

1.2.3.2) Standardised Uptake Values

The Standardised Uptake Value (SUV), also referred to as the Differential Uptake Ratio (DUR) or Dose Absorption Ratio (DAR), is the most popular method of quantification and is expressed as:

$$SUV_{BW} = \frac{Q \times W}{Q_{inj}} \qquad [1.3]$$

where Q is radiotracer concentration (MBq/l), Q_{inj} is injected activity (MBq) and W is body weight (kg) (Woodard *et al*., 1975).

SUVs can either be taken as the mean value (SUV_{mean}), or the SUV_{max} in a region or volume of interest (ROI/VOI) of the tumour, segmented manually or by automatic methods such as thresholding. SUV_{max} is operator independent and will almost always have the same value regardless of the technique used to draw the VOI. However, random errors can distort individual pixels so values may not be a good representation of the tumour as a whole. Due to this, SUV_{mean} is used in some investigations but this causes problems as segmentation of the VOI by the operator is very subjective and the value will depend on the VOI selected (Berkowitz *et al.*, 2008). Therefore, SUV_{max} is usually used clinically, and in most studies, although SUV_{mean} values are often recorded too.

The PERCIST criteria suggest the use of the SUV_{peak} parameter which aims to reduce the potential distortion of SUV_{max} by taking a VOI around it. It has been thought to be a more robust parameter to noise than SUV_{max} (Lodge *et al*., 2012). However, there are issues in defining the small volume around the SUV_{max} as the method for doing so can vary the result considerably (Vanderhoek *et al*., 2012). A typical, normal tissue SUV is ~1. However, some organs including the brain and liver actively take up more FDG and will therefore have SUVs greater than this (Hallett *et al*., 2001). Equally, tissues including the lung and adipose tissue have less of a need for glucose and so have SUVs of <1, while most cancers will have SUV >1. A SUV of ~2.5 has been commonly used as a cut-off point to separate benign and malignant processes (Hicks, 2005).

SUV does not require any blood sampling and can give values that can be used to aid diagnosis. As a result of not requiring blood sampling, SUV needs the injected dose to be calibrated and

measured accurately and makes the assumption that plasma clearance will remain the same, which is not always the case. This is a particular issue when comparing pre and post therapy SUVs as plasma clearance of FDG may change after therapy due to varying uptake in other tissues potentially affecting the relationship between uptake at a specific time and the administered dose (Lammertsma *et al*., 2006). SUV has been claimed to be a flawed quantitative method as it is subjected to many sources of variability, which is particularly an issue when comparing SUVs over different institutes where there can be variability of up to 25% (Fahey *et al*., 2010). However, SUVs are highly reproducible on the same scanner and using the same protocols (Nahmias and Wahl, 2008), although there is more of an issue in tumours with low ^{18}F-FDG uptake (de Langen *et al*., 2012). The most notable sources of variability are patient size, measurement times i.e. time from injection of the tracer to the start of the scan, plasma glucose levels, the partial volume (PV) effect and the position of the VOI. These potential limitations to quantification are discussed further in 1.2.3.4 to 1.2.3.8.

1.2.3.3) Volumetric Measures

The use of volumetric measures for assessing response to therapy has become popular over the last few years (Roedl *et al*., 2009; Everaert *et al*., 2011; Gulec *et al*., 2011). Using a volumetric parameter combining TV with the intensity of uptake within it was first introduced by Larson *et al*. (1999) and defined as total lesion glycolysis (TLG) (Larson *et al*., 1999). The TLG for each tumour can be added together to calculate a TLG value for the disease. The change between these parameters between pre- and post- therapy scans can be used to measure response and is defined as the δTLG or the Larson-Ginsberg Index (LGI):

$$\delta TLG = \frac{(SUVmean_1 * Volume_1) - (SUVmean_2 * Volume_2)}{(SUVmean_1 * Volume_1)} * 100 \quad [1.4]$$

where 1 and 2 are pre- and post- therapy scans, respectively.

Other studies have used very similar measures of TV and SUV_{mean}, differing only by segmentation methodology of TV, subtle corrections and terminology. The TLG has also been termed the effective glycolytic volume (EGV), total glycolytic volume (TGV) and metabolic tumour burden (MTB) or metabolic burden (MB) (Nakamoto *et al*., 2002; Francis *et al*., 2007; Berkowitz *et al*., 2008). Methodology can change slightly between studies. For example, while Larson *et al*. (1999) used a fixed ROI placed by an observer on the PET image to define TV (Larson *et al*., 1999), Berkowitz *et al*. (2008) used a manual segmentation on the CT and used the recovery coefficient (RC) to correct for the PV effect (Berkowitz *et al*., 2008), calculating MB as:

$$MB = \frac{SUV_{mean} * CT\ (V_{CT})}{RC} \qquad [1.5]$$

where V_{CT} is the segmented CT volume and SUV_{mean} CT is the SUV_{mean} within the CT volume on the PET image.

1.2.3.4) Limitations of SUV due to Scanner Issues

SUV has become the most used semi-quantitative measure because of its normalisation and simplicity. However, it has a number of limitations which can cause variation in values. With regards to the PET scanner itself, noise and image resolution have been shown to have a substantial impact (up to 50%) on SUVs in phantom simulations, however, SUV ratios used for response monitoring are not affected as much, implying SUVs may be better equipped to deal with identifying response (Boellaard *et al*., 2004).

Differences in SUV of up to 40% between high (7mm) and low (10mm) resolution images have been reported (Westerterp *et al*., 2007), corresponding with similar results when reconstructing data with less iterations than normal, although there has been no significant differences found

between measured and segmented AC methods (Jaskowiak *et al*., 2005). Significant differences in SUV were found when using different numbers of iterations and subsets in OSEM and FBP reconstruction algorithms (Visvikis *et al*., 2001; Ivanovic *et al*., 2004), although no statistical differences between SUVs on OSEM and FBP reconstructed images have been reported (Krak *et al*., 2005). The effect of the PET/CT hardware used and the positioning of the subject has little effect on SUVs (Doot *et al*., 2007), however, sufficient calibration of the PET scanner should be conducted before obtaining images for quantification to establish accuracy of at least 10% (Geworski *et al*., 2002).

1.2.3.5) Limitations of SUV due to Physiology

Body weight is commonly used to normalise body concentration for SUVs, however, it can be inaccurate because it does not give a specific volume and can be variable. This is particularly true in patients with a high body mass index (BMI) who have higher SUVs than they should because of the low uptake of FDG in adipose tissue. Due to this, other normalisation methods have been suggested. SUL (or SUV-lean or SUV_{LBM}) replaces body weight with lean body mass (LBM) for normalising SUV (Zasadny and Wahl, 1993). Similarly, SUV_{BSA} replaces body weight with body surface area (BSA) for normalising SUV (Kim *et al*., 1994), and has been found to correlate with kinetic modelling methods better than SUV normalised by body weight (Graham *et al*., 2000).

However, a study has shown that the method of estimating LBM for SUVs can cause substantial error in comparison to accurate dual-energy X-ray absorptiometry (DEXA) (Erselcan *et al*., 2002). In addition, these methods only improve characteristics of SUVs over the population as a whole, as identifying abnormal uptake requires estimation of normal values which depend entirely on the patients in the study (Hallett *et al*., 2004). A weight index (SUV_m) has been

proposed to reduce variability where the coefficient of variation of SUVs was below a third (Thie *et al*., 2007). This approach means population averages are dimensionless unlike SUV_{LBM} or SUV_{BSA}.

The amount of blood glucose is another psychological factor which can cause variation in SUVs as SUV can be reduced if there is more glucose to compete with FDG. This has been studied in head and neck cancer patients and in bronchial carcinomas (Langen *et al.*, 1993; Lindholm *et al*., 1993). Both investigations found differences in SUVs and this has led to corrections for blood glucose (SUV x glucose) being used in calculating SUVs, however, these have not always led to better results (Diederichs *et al*., 1998; Hallett *et al.*, 2001). There are conflicting results with regards to the effect of glucose on SUV. One study has shown that SUVs for lung tumours in diabetic patients (with elevated glucose levels) were not significantly different to those from non-diabetic patients (Gorenberg *et al*., 2002). However, it has been suggested that this may be due to differences in tumour type and glucose transporter expression (Hallett, 2004). Generally, none of the correction methods mentioned have been fully validated and a study using different methods of correcting SUV for blood glucose, tumour size and BSA/LBM in lung nodules showed no improvement in accuracy compared to traditionally calculated SUV (Menda *et al*., 2001).

Other physiological factors which can affect SUV include patient breathing or motion and inflammatory processes (Erdi *et al.*, 2004; Boellaard *et al*., 2008). Patient breathing and motion can cause resolution loss and artefacts and when imaging the lungs, changes in motion of 10mm have been found to cause SUV changes of 20% (Erdi *et al.*, 2004). SUVs can be affected by tense patients and uncomfortable conditions which can cause increased FDG uptake in muscle and brown fat which will have an effect on the image, and therefore the SUV (Boellaard *et al*.,

2008). Finally, inflammatory processes can also cause FDG uptake, and if this occurs around tumour areas there can be false positive results (Boellaard *et al*., 2008).

1.2.3.6) Limitations of SUV due to Partial Volume Effect

The PV effect describes what happens when signal intensities mix near the boundaries of tissues causing loss of intensity in small objects such as tumours (Figure 1.4). The effect arises from limited spatial resolution which can produce inaccurate and blurred images. This is of a particular problem in PET studies where resolution is not as high as other modalities including CT or MRI. The PV effect has been shown to underestimate small lung lesion phantoms by up to 91% (Feuardent *et al*., 2005).

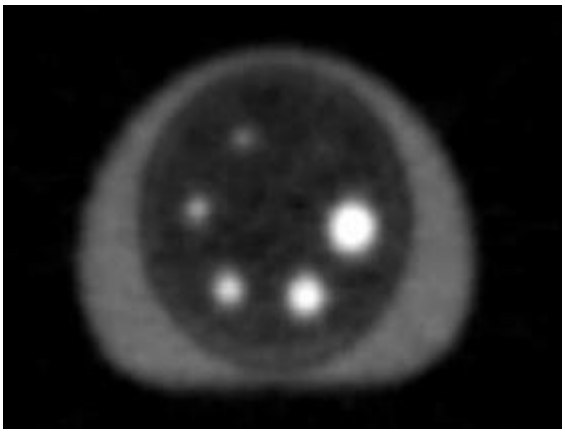


Figure 1.4: Partial Volume Effect in a Phantom Image

The phantom contains a number of cylinders of the same activity but the measured activity decreases depending on the size of the cylinder due to the PV effect (Taken from Barrington, 2006).

PV corrected images can give better T/B ratios and improve quantitative image analysis. There have been many differing approaches to PV correction including noise models and geometric transform matrix (GTM) models (Rousset *et al.*, 1998; Aston *et al.*, 2002; Frouin *et al.*, 2002), which incorporate popular use of high resolution magnetic resonance (MR) images in

combination with the known point spread function (PSF) of the imaging system (Bencherif *et al.*, 2004; Quarantelli *et al.*, 2004).

Kessler *et al.* (1984) introduced the RC, a ratio of observed intensity to true intensity that can be seen in an image (Kessler *et al*., 1984). This ratio can be used to boost observed intensities up to the true values and has been used in a number of studies (Uemura *et al*., 2000; Asselin *et al*., 2004; Teo *et al*., 2007). It has been stated that PV correction of SUV produces a more accurate assessment of disease activity and will help in prognosis (Basu and Alavi, 2007), however, until an accepted method of PV correction is available care must be taken to standardise acquisition and analysis (Soret *et al.*, 2007). Although correction for PV may improve the estimation of uptake, most techniques are so sensitive to object size, shape and heterogeneity there is a limit to how much they can achieve (Hallett, 2004). For any object smaller than about two times the full width half maximum (FWHM), typically less than 15-20mm, there is no accurate and precise method for correction available (Boellaard *et al*, 2008).

SUVs of small lesions (<2cm) have been corrected for the PV effect by using CT data to measure the lesion and calculating a corrected SUV using this lesion size (Hickeson *et al*., 2002). This is defined as:

$$\text{corSUV} = \frac{\left(\dfrac{\text{Region Activity (MBq)} - \text{Background Activity (MBq)}}{\text{Lesion's size on CT scan (cm}^3\text{)}}\right)}{\left(\dfrac{\text{Injected dose (MBq)}}{\text{Body Weight (g)}}\right)} \qquad [1.6]$$

where background activity = activity / volume in background * (region's volume – lesion's size on CT cm^3). This calculation of SUV corrects for the understimation of true metabolic activity of

a lesion caused by resolution (by subtracting background activity) and partial volume effects (by using the lesion size on CT). The problem with this technique is it assumes that anatomical CT TV and PET MTV are the same and this is not always the case.

1.2.3.7) Limitations of SUV due to Imaging Time

The time between the injection of FDG and the start of the scan can cause variation in SUV as concentrations of FDG take time to reach a plateau before imaging. In patients with lung cancer, an approximate 60min wait was found to underestimate SUV significantly in most tumours, as FDG uptake continues past 60min and does not reach 95% plateau until, on average, 298min pre-treatment and 154min post-treatment (Hamberg *et al.*, 1994). These findings have been supported by further evidence showing increases in SUV of up to 30% and greater tumour detectability depending on the time delay between FDG injection and the start of scanning (Hustinx *et al*., 1999; Lodge *et al*., 1999; Stahl *et al*., 2004). Although these findings suggest a longer waiting time before scanning would be beneficially, practically it is usually not feasible. An imaging time of between 50-70min after injection is considered both optimal and practical by the EORTC, PERCIST and other studies (Lowe *et al*., 1995; Young *et al*., 1999; Wahl *et al*., 2009).

1.2.3.8) Other Limitations of SUV

SUVs can be affected by inaccurate measurement of FDG administration. Errors can arise due to inaccurate cross-calibration between the PET scanner and dose calibrator for measuring patient dose, unaccounted for residual activity in the syringe (or other administration system), and the use of injection time in calculations instead of dose calibration time (Boellaard *et al*., 2008). SUVs can also be affected by respiratory motion (Goerres *et al*., 2002a), the presence of metal

implants or prostheses (Goerres *et al*., 2002b; Goerres *et al*., 2003b; Kamel *et al*., 2003), and the use of CT contrast agents such as Iodine and Barium (Antoch *et al*., 2003; Cohade *et al*., 2003). All can cause artefacts and over- or under- estimate tracer concentration activity, affecting SUVs by ~5% in regard to contrast agents (Dizendorf *et al*., 2003). Respiration artefacts can be limited by using breathing protocols, such as free breathing and normal expiration, during image acquisition but these do not produce perfect results (Goerres *et al*., 2003a).

One of the major problems with SUVs is that comparing values from different institutes is problematic because of all the limitations noted. Any cut-off values designed to estimate, for example, whether a tumour is malignant or benign are often institute specific, particularly if they are affected by differences in acquisition protocols, reconstruction algorithms and ROI definition (Lammertsma *et al*., 2006). This is one of the reasons why standardisation is required, to minimise fluctuations so that semi-quantitative analysis techniques can be used across institutions in MCTs. It is important to note that despite these limitations SUVs are still reliable and reproducible within an institute, as has been shown in a number of studies (Minn *et al*., 1995; Nakamoto *et al*., 2002; Paquet *et al*., 2004).

1.2.4) Full Quantitative Methods using Kinetic Models

1.2.4.1) Kinetic Modelling

Kinetic modelling links measured activity levels in functional scans with physiological parameters such as the metabolism of glucose by tumours, organs, and tissues, mathematically describing the movement of FDG in cells (Basu *et al*., 2007). Modelling was first used to calculate glucose utilisation in the brain (Sokoloff *et al*., 1977; Reivich *et al*., 1979; Phelps *et al*., 1979). FDG enters glucose-consuming cells *via* glucose transporters (GLUTs) where it is

phosphorylated by hexokinases. Unless dephosphorylation occurs, phosphorylated FDG remains trapped in the cell (Castell and Cook, 2008).

1.2.4.2) Nonlinear Regression

Nonlinear Regression (NLR) analysis is the most complex method of quantification used in PET studies. It has been adapted for use outside of the brain to evaluate glucose metabolism in specific lesions, for example, in liver tumours (Okazumi *et al.*, 1992). A generalised version of the compartmental model used to achieve this is shown in Figure 1.5.

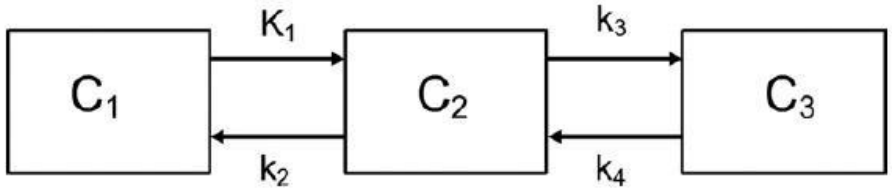


Figure 1.5: Three Compartmental Model of FDG Behaviour

The model uses rate constants to evaluate glucose metabolism (Taken from Basu *et al*., 2007).

NLR uses an algorithm which fits values to rate constants, typically by nonlinear least squares, using a two or three compartmental model with an arterial plasma input to measure the net influx rate constant for FDG (Castell and Cook, 2008). This can be defined as:

$$MR_{glu} = \frac{Cp}{LC} \times \frac{K_1 * k_3}{k_2 + k_3} = \frac{Cp}{LC} \times Ki \qquad [1.7]$$

where MR_{glu} is the glucose metabolic rate (MR_{glu}), Cp is the plasma glucose concentration, K_1 and k_2 are rate constants for forward and reverse transport of FDG respectively, K_i is the net rate of influx and LC is the lumped constant (Basu *et al*., 2007). There are now many different versions of this technique, from even more complex methods involving two ROIs and six

compartments (Wu *et al*., 1995), to time saving variants including the simplified kinetic method (SKM) which requires just one venous blood sample (Hunter *et al*., 1996), and the Sadato method which estimates Ki based on its relationship with SUV (Sadato *et al*., 1998). The most widely used simplified method is Patlak analysis.

1.2.4.3) Patlak Analysis

Patlak analysis is a simplified method of NLR and was originally developed for evaluating transfer constants from blood to brain (Patlak *et al*., 1984). It is a linearisation of the NLR model and still needs dynamic scanning and an arterial input function (AIF) for worthwhile results, but is a more rapid and robust method and produces results which correlate well with NLR (Cheebsumon *et al*., 2011). Simplifications mean there is no separation between GLUTs and hexokinases, and dephosphroylation is assumed to be negligible, making it less accurate in comparison to NLR. However, Patlak analysis does allow the possibility of generating functional images of metabolism using pixel level calculations (Lammertsma, 2001).

Less invasive methods of blood sampling have been suggested for Patlak analysis. A population-based arterial blood curve using data from 10 patients has been developed (Takikawa *et al*., 1993), as has image derived input functions (IDIFs) for non-invasive quantification of the cerebral metabolic rate (Chen *et al*., 1998). Both techniques show good correlation with real arterial input, making them worthwhile alternatives but further validation is required and they will still lack the accuracy of blood sampling. Patlak separates out metabolised and unmetabolised FDG in blood and intracellular spaces of the cell and by using arterial sampling, the integral under the AIF can be used for more accurate normalisation in compared to SUV (Freedman *et al*., 2003).

1.2.4.4) Limitations of Kinetic Modelling

While NLR provides one of the most accurate estimates of tumour glucose use, it comes at the cost of time and inconvenience as it needs arterial blood sampling (Dimitrakopoulou-Strauss *et al*., 2002). Due to this, it is rarely used in clinical settings as absolute quantification is not usually necessary and limited statistical data can lead to errors in fitted parameters. More simplified modelling approaches, such as Patlak analysis, also have similar issues (Hallett, 2004). Although there is a need for these techniques, the complexity and time taken to conduct them makes it impractical in a clinical context. As a result, these techniques are not explored further as the clinical data that will be investigated are static scans with no arterial sampling so kinetic modelling would not be possible. Therefore, the main focus will be on SUVs and other semi-quantitative methods, such as volumetric measures.

1.2.5) Comparison of Quantitative Methods

Quantitative methods have correlated well with each other in a number of studies with NLR being used as the gold standard (Graham *et al*., 2000; Wu *et al*., 2001; Hoekstra *et al*., 2002; Kroep *et al*., 2003; Krak *et al*., 2003; Brenner *et al*., 2004). SUVs are the simplest form of quantitative measurement but results suggest that despite their limitations they are reliable and can be used successfully. While NLR and Patlak analysis provide greater accuracy and have the ability to give exact biological values for specific processes, SUVs offer a reliable method of identifying response to therapy in a clinical environment where kinetic modelling is not possible. Despite positive correlations, discrepancies can be found between methods and these must be considered (Freedman *et al*., 2003).

1.2.6) Imaging Recommendations for Response Assessment and Quantification

To produce accurate and reproducible data a number of reports have described standardised methodology and protocols for PET scanning for accurate quantification of images and identifying response to therapy. They include the EORTC guidelines (Young *et al*., 1999), National Cancer Institute (NCI) consensus recommendations (Shankar *et al*., 2006), Procedure Guideline for Tumor Imaging 1.0 (Delbeke *et al*., 2006), PERCIST (Wahl *et al*., 2009), and European Association of Nuclear Medicine (EANM) procedure guidelines for tumour imaging (Boellaard *et al*., 2010), the latter being heavily drawn from the Netherlands protocol (Boellaard *et al*., 2008).

There are differences between the guidelines but they mostly agree on specific standards in PET imaging (Table 1.3). These include: (i) Patients should be fasting for at least 4-6h before a PET scan, (ii) patients should be well hydrated, (iii) patients should not have done any strenuous exercise before the scan, and (iv) plasma glucose levels should be <200mg/dl. Before the scan, patients should be resting in a warm, dimly lit, quiet room and should urinate before the start of the scan. There should be ~1h wait between the injection of ^{18}F-FDG and the start of the scan with this time gap made standard for all scans. The timing of the post-treatment scan should be between 10-14 days after the start of treatment. The aspect of the guidelines which differs most is the way images should be analysed, with varying standards suggested in terms of correction of SUVs, response measurements and how TV should be obtained.

Overall, the guidelines are very similar and are designed to get the most reproducible ^{18}F-FDG uptake and SUV measurements possible, preventing variations between scans. This is of particular importance in MCTs where camera calibration, image reconstruction and data analysis/settings can mean a variability of more than 50% on SUVs (Boellaard *et al*., 2009).

Changes in SUV of up to 25% are possible due to instrument and analysis factors from different centres, however, this is reducible by using a central reading from one institute and performing rigorous quality control procedures (Fahey *et al*., 2010). Inter-institution calibration and a standardised scanning methodology along with strict quality control measures are vital to the success of MCTs and strict standardisation of PET imaging is of the utmost importance (Boellaard *et al*., 2010).

Currently, improvement in this area is needed as recent reports by eight imaging response assessment teams, funded by the NCI, showed there were major areas of variation in ^{18}F-FDG dose, uptake time, handling of diabetic patients, duration of fasting and acquisition protocols used (Beyer *et al*., 2011; Graham *et al*., 2011). This is why a standardised and globally accepted procedure of ^{18}F-FDG PET/CT imaging is needed in all quantitative studies both in research and clinical, diagnostic practice (Boellaard *et al*., 2011). Factors affecting quantifying tracer uptake and ways to correct issues such as calibration errors are being investigated (Doot *et al*., 2010; Lockhart *et al*., 2011).

Guideline	EORTC (1999)	NCI (2006)	Tumor Imaging 1.0 (2006)	PERCIST (2009)	EANM (2010)
Fasting	6h	4h	4-6h	4-6h	6h
Hydration	500ml after injection	500ml after injection	Encouraged	N/A	1 litre in 2h before injection
Exercise	N/A	None for 24h before scan	N/A	N/A	None for 6h before scan
GL Expected	4-7 mmol/l 72-126mg/dl	<120mg/dl	<150mg/dl	<200mg/dl	<120mg/dl
GL Cancel	N/A	>200mg/dl	150-200 mg/dl	>200mg/dl	>120mg/dl
Preparation	Relaxation encouraged	Patients should be in a warm, comfortable, dimly lit, quiet room	Patients should be seated or recumbent in a quiet, dimly lit room and should urinate before the scan	N/A	Patients should be in a warm quiet, darkened room, refrain from talking and urinate before the scan
Scan Start Time from Injection	N/A	60min (+/- 10min)	At least 45min. Between 60 – 90min	50 – 70min (+/- 15min)	60min (+/- 5min)
Treatment to Scan Time	Within 14d	~14d	N/A	>10d	>10d
Image Analysis	Should use SUV_{BSA} and manual delineation of tumours	No single optimal method of analysis is suggested but use of SUV_{LBM} or SUV_{BSA} is recommended	Use of SUV_{LBM} or SUV_{BSA} is recommended	Recommends use of SUL, SUL_{peak}, and TLG using a background based segmentation	Mentions the use of all SUV corrections and recommends SUV_{peak} and delineation using thresholding

Table 1.3: Comparison of Guidelines for ^{18}F-FDG PET Response Assessment

The five guidelines compared are (i) EORTC criteria (Young *et al*., 1999), (ii) Consensus Recommendations in NCI trials (Shankar *et al*., 2006), (iii) Procedure Guideline for Tumor Imaging 1.0 (Delbeke *et al*., 2006), (iv) PERCIST (Wahl *et al.*, 2009), and (v) EANM Procedure Guidelines for Tumour Imaging (Boellaard *et al*., 2010). Specific guidelines compared are (a) minimum amount of time a patient should fast before a scan, (b) recommendations for the hydration and exercise of the patient before a scan, (c) the plasma glucose levels that are expected and the maximum level over which a scan would be cancelled, (d) guidelines for patient preparation, (e) the scanning start time from ^{18}F-FDG injection, (f) the time between the first treatment of the patient and the post-therapy scan, and (g) guidelines for image analysis. GL = Glucose Level. N/A = Not Applicable. SUV_{LBM} = SUL.

1.3) Image Analysis for Identifying Response to Therapy

1.3.1) Overview of Image Analysis for Identifying Response

Identifying response to therapy can involve many methodologies for categorising and scoring response between pre- and post- therapy images. Qualitative techniques, including visual analysis by trained consultants, are predominantly used in the clinical environment but there is a great deal of interest and research in quantitative methods (Tomasi *et al*., 2012; Lambin *et al*., 2012). Image analysis techniques can produce quantitative measurements, including SUVs and TVs, which could be of great benefit in early response assessment and in MCTs where qualitative methods are flawed due to a lack of comparability and reproducibility (Tomasi *et al*., 2012). Image analysis can come in many different forms but in terms of medical imaging, and PET studies in particular, some of the main tools are the ability to register images, segment tumours and extract intensity data such as SUVs and histograms. Other techniques such as quantifying texture and shape in an image by analysing the spatial distribution of voxels and their intensities could also be useful in identifying response in pre- and post- therapy PET/CT studies. While there has been a great deal of research in the use of SUVs, assessing response to therapy using TV, TLG, intensity volume histogram (IVH) parameters and texture analysis is gaining a lot of interest in research (Tomasi and Rosso, 2012). These methods may provide more information, related to the size, variation, and pattern of intensity within a tumour rather than just a maximum value, and this is very appealing for assessing response.

1.3.2) Image Registration

The goal of image registration is to match two images so that corresponding features are aligned with one another. This can mean registering different patients, different modalities, scans at different times, including differences in motion between scans within a few minutes of each other, or a combination of all three. To achieve this, a spatial transform is obtained which maps

the points on one image to those on another. Registration techniques can be classified in various ways. Rigid registration describes techniques which involve the rotation and translation of one image to match a target image, while non-rigid registration not only allows rotation and translation but also stretching and deformation of images in order to match them. This is often needed due to image acquisition or biological differences, including differing image protocols and patient positioning (Crum *et al*., 2004). A registration algorithm is usually made up of three components. A similarity measure which dictates how well the two images are matched, a transformation model which specifies how the source image is changed to match the target and an optimisation process which varies transformation parameters in order to get the best registration (Crum *et al*., 2004). Registration is rarely used routinely in a clinical environment but is widely used in research, particularly rigid algorithms for registering brain images.

Registration of PET/CT studies is occasionally done clinically to allow a clinician to align pre- and post- therapy scans with each other to aid visual analysis. This is usually done using simple manual registration or by a computationally efficient rigid algorithm as precise and accurate registration is not of critical importance in this scenario. For more quantitative purposes, more accurate registration is required and different techniques are applied. For assessing response to therapy, the most interesting prospect for registered pre- and post- therapy PET/CT images is the use of parametric or subtraction images to characterise tumour change. This concept was first investigated in 15 patients with lung cancer and has since been used to successfully predict RECIST-based response on CT scans performed 5-8 weeks after initial post-therapy PET/CT scans in 28 patients with colorectal cancer (Necib *et al*., 2008; Necib *et al*., 2011). Parametric images were obtained by rigidly registering CT images and applying the transformation to the PET component of the PET/CT scans. Bi-parametric graphs of tumour voxels in the pre- therapy PET scan can be plotted against the voxels of the subtracted PET image (pre-therapy – post-

therapy) (Figure 1.6). It is claimed that the volume affected by changes in intensity and mean change in SUV within this volume can be used to identify response to therapy.

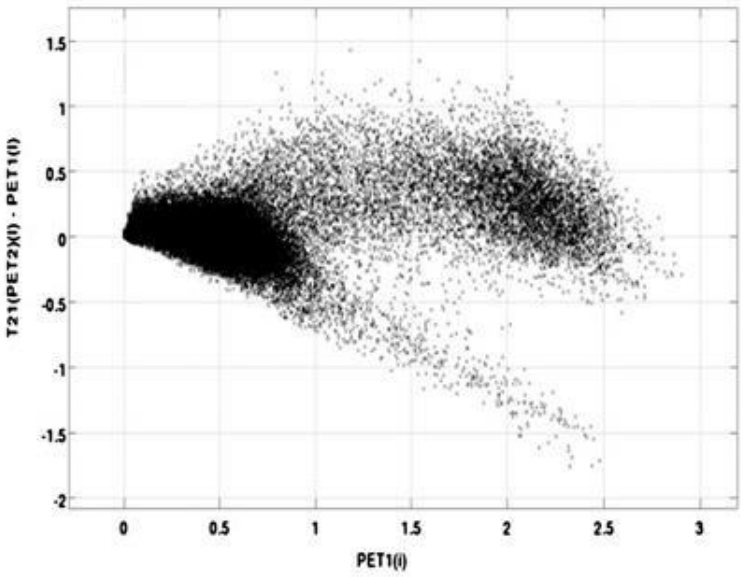


Figure 1.6: Bi-parametric Graph using a Subtraction Image

The graph plots voxels from a liver tumour on the subtraction image (pre-therapy PET – post-therapy PET) with the pre-therapy image to see if changes in intensities can be identified (Taken from Necib *et al*., 2011).

1.3.3) Tumour Volume and Total Lesion Glycolysis

TV and TLG are being increasingly used in quantifying response in pre- and post- therapy PET scans (Francis *et al*., 2007; Berkowitz *et al*., 2008; Benz *et al*., 2008; Roedl *et al*., 2009; Everaert *et al*., 2011; Gulec *et al*., 2011). While the basis of these metabolic measurements remain the same, methodology changes slightly depending on the type of cancer investigated, correction factors used and, principally, the segmentation method used for obtaining TV. Table 1.4 illustrates the differing use of TLG, and its many other terminologies, in a number of PET studies using TV and TLG.

Study	Type of Cancer	Term Used	Segmentation(s) Used
Larson *et al*. (1999)	Lung, rectal, esophageal and gastric	TLG	40% of SUV_{max}
Akhurst *et al*. (2000)	Renal	TLG	Manual Segmentation
Bural *et al*. (2006)	Atherosclerosis in the aorta	Atherosclerotic burden / MTB	Manual Segmentation
Francis *et al*. (2007)	Mesothelioma	TGV	GRAB algorithm - uses SUV_{mean} and BG based threshold
Benz *et al*. (2008)	Soft tissue Sarcoma	TLG	CT segmentation
Berkowitz *et al*. (2008)	NHL	MTB	CT segmentation or 40% of SUV_{max}
Boucek *et al*. (2008)	Phantom Study	TGV	50% of SUV_{max}, BG adapted and GRAB algorithm
Costelloe *et al*. (2009)	Osteosarcoma	TLG	Manual Segmentation
Roedl *et al*. (2009)	Esophageal adenocarcinoma	TLG	Fixed 2.5 SUV threshold
Lee *et al*. (2010)	Mesothelioma	TLG	PERCIST criteria: liver BG + 2 S.D
Nowak *et al*. (2010)	Mesothelioma	TGV	GRAB algorithm
Arslan *et al*. (2011)	Small cell lung cancer	TLG	Fixed 2.5 SUV and 50% of SUV_{max}
Everaert *et al*. (2011)	Rectal adenocarcinoma	TGV	2.5 SUV threshold in Manual Segmentation
Gulec *et al*. (2011)	Colorectal cancer in liver metastases	TLG	Threshold of liver SUV_{max}
Hatt *et al*. (2011)	Oesophageal	TLG	FLAB probability method and adaptive threshold.
Abd El-Hafez *et al*. (2012)	Oral cavity squamous cell carcinoma	TLG	Fixed 3 SUV threshold
Chen *et al*. (2012)	NSCLC	TLG	50% of SUV_{max}
Chung *et al*. (2012)	Ovarian	TLG	40% of SUV_{max}
Dibble *et al*. (2012)	Oropharyngeal squamous cell carcinoma	TGA	Gradient-based method and 38%, 50% and 60% of SUV_{max}
Lim *et al*. (2012)		TLG	42% of SUV_{max}
Im *et al*. (2012)	Osteosarcoma	TLG	Various fixed thresholds of SUV
Liao *et al*. (2012)	NSCLC	TLG	Gradient-based method
Sharma *et al*. (2012)	Lymphoma (Paediatrics)	MTB	% of SUV_{max} dependent on CT tumour length
Zhang *et al*. (2012)	NSCLC	TLG	Gradient-based method

Table 1.4: Comparison of Literature using Tumour Volume and Total Lesion Glycolysis in PET

Comparison of PET studies investigating TV and TLG showing different terminology, segmentation methods and types of patients. All definitions of TLG, TGV, MTB and total glycolytic activity (TGA) were calculated as TV * SUV_{mean} within TV. The exception being a study by Larson *et al*. (1999) in which SUV_{mean} was taken as a large 2-D ROI around the SUV_{max} rather than the whole segmented TV (Larson *et al*., 1999). Segmentation methods are discussed in 1.3.4. BG = Background.

Table 1.4 includes some studies using TV and TLG on just pre-therapy scans rather than a comparison between pre- and post- therapy scans but the methodology for obtaining the measures is the same. Studies used typical SUV measurements with a few exceptions which used SUL (Costelloe *et al*., 2009), SUV corrected for plasma glucose levels (Larson *et al*., 1999), and the use of RC on small tumours (Berkowitz *et al*., 2008). Success of using TV and TLG measures to predict response to therapy has been mixed. While some studies show they perform more favourably than measures such as SUV_{max} (Francis *et al*., 2007; Berkowitz *et al*., 2008; Roedl *et al*., 2009; Hyun *et al*., 2010; Lee *et al*., 2010), others have found the reverse (Benz *et al*., 2008; Costelloe *et al*., 2009; Everaert *et al*., 2011). There is enough evidence of success to suggest these methods require further investigation and differing results are likely due to different types of cancer and segmentation methods, the latter being an increasingly prominent area of image analysis in PET studies in its own right. As the terms TLG, TGV, TGA, and MTB all refer to the same calculation of the product of TV and SUV_{mean} within the TV, to avoid confusion this measurement will be defined as TLG.

1.3.4) Image Segmentation

The aim of segmentation is to accurately delineate a particular object from the rest of the image whether this is an anatomical landmark on an MRI image, a tumour on a CT image or high areas of uptake on a PET image. The segmentation of disease on a PET image is needed to assess TV and TLG and is also necessary for other methods of analysis. An obvious method of segmentation, used often in radiotherapy planning, is visual contouring of a tumour on individual 2-D slices, over a 3-D image, by an experienced physician (Avril *et al*., 1997). The main limitation with this technique is there is a degree of observer variability, as each physician will segment lesions differently. Many user-defined protocols have been constructed to try to reduce this variability with some success (MacManus *et al*., 2007). However, observer variability will always be an issue when using user defined contouring to segment tumours on

PET images and this had led to the development of a number of semi-automated and automated methods for segmenting PET lesions.

Semi-automated threshold contouring allows the segmentation of a PET lesion based on a threshold, whereby all voxels in the tumour area are included if they are above the threshold value. Typical region growing algorithms would 'grow' from a user defined starting point placed somewhere in the tumour and include all connected voxels above the threshold to create a TV. There are a number of ways to choose a threshold including taking a percentage of the SUV_{max} (Erdi *et al*., 1997), using a fixed SUV (Paulino and Johnstone, 2004), or by using a calculation which takes into account background intensity in the image (Boellaard *et al*., 2004). Using a fixed SUV as a cut off for segmentation is useful for many reasons. Its simplicity makes it easy to implement and for physicians to understand what has been segmented and an SUV of 2.5 has been shown to be useful for separating benign and malignant tumour (Paulino and Johnstone, 2004). It should be noted that some disease can have a SUV of <2.5 while many active inflammatory processes can have SUVs higher than this (Hicks, 2005; Kinahan and Fletcher, 2010). The main limitation with using a fixed SUV threshold is that there are no 'normal' SUVs as they are too easily affected by bias and so using a fixed value, particularly from different scanners or centres, is unrealistic as SUVs are likely to be inconsistent (Boellaard, 2009).

Taking a percentage of the SUV_{max} does not face the same problem as using a fixed SUV, in that it does not rely on the same SUV for all tumours/images and, therefore, does not suffer from a dependence on consistent SUVs, as there can be significant variation (Boellaard *et al*., 2004). It does suffer from the fact that the optimal percentage is often so different for each individual lesion on each image that a fixed percentage is unlikely to be the best threshold to obtain the correct SUV for segmenting the tumour. Different percentages of SUV_{max} for tumour segmentation are often used for different studies (Erdi *et al*., 1997; Boellaard *et al*., 2004;

Boucek *et al*., 2008; Dibble *et al*., 2012), the percentage used seemingly dependent on the data in question and the author's discretion. It is acknowledged there is no one appropriate percentage of SUV_{max} which can be used as an optimal threshold (Biehl *et al*., 2006). Rather than use SUV_{max}, region growing techniques can use SUV_{mean} of a region to calculate a fixed threshold (Black *et al*., 2004), or continually adapt the threshold to a percentage of SUV_{mean} as the region grows (Green *et al*., 2008). These methods eradicate the problem of finding an optimal threshold % of SUV_{max} but only take into account the tumour and not the intensity of the background around it.

The source-to-background (S/B) ratio in PET images can be an important factor in obtaining the correct TV and this has lead to segmentation methods which use a background VOI rather than tumour values (Zasadny *et al*., 1998; Wahl *et al*., 2009; Gulec *et al*., 2011). PERCIST guidelines suggest segmentation should be based on a threshold equal to normal-liver mean, based on a ROI of 3cm in diameter, plus two standard deviations (S.D.) (Wahl *et al*., 2009). This calculation was originally proposed by Zasadny *et al*. (1998), adding three S.D. to the mean rather than two. Another study investigating TLG measurements used a background value equal to the SUV_{max} in the liver to segment disease (Gulec *et al*., 2011). The combination of background and SUV_{max} to determine the threshold for segmentation has been used with a threshold based on 50% of the background and SUV_{max} (Boellaard *et al*., 2004).

Many methods have investigated the relationship between tumour SUV, background SUV, T/B ratios and tumour size (Daisne *et al*., 2003; Yaremko *et al*., 2005; Davis *et al*., 2006; Drever *et al*., 2006; Van Dalen *et al*., 2007; Jentzen *et al*., 2007; Nehmeh *et al*., 2009). Two of these algorithms use a threshold based on the background added to a certain percentage of a SUV_{max} background corrected value (SUV_{max} – background), where the percentage is determined by the contrast level and/or tumour size based on phantom experiments (Davis *et al*., 2006; Drever *et*

al., 2006). Other methods use a similarly constructed methodology and formula but continually adapt the percentage or relative threshold level (RTL) to iteratively calculate the correct RTL until it does not deviate significantly from the last iteration (Van Dalen *et al*., 2007; Jentzen *et* al., 2007; Nehmeh *et al*., 2009). The issue with these types of methods is by using phantom measurements and parameters specific to a particular PET scanner, the methodology is not easily transferable from one PET system to another (Lee, 2010).

Similar methods have also used SUV_{mean} of a tumour rather than SUV_{max} along with background to define a threshold. Two studies investigating the relationship between the optimal threshold and tumour SUV_{mean} and background came up with similar calculations for a threshold for segmentation based on the SUV_{mean}, background and constant variables (Nestle *et al*., 2005; Nestle *et al*., 2007). Additionally, an adaptive region growing segmentation method has been described which uses both SUV_{mean} and background to segment PET tumours (Boucek *et al*., 2008), a modification of the algorithm using just SUV_{mean} (Green *et al*., 2008). The disadvantage of using background uptake in segmentation methods is the background region is likely to vary depending on how the VOI is obtained, particularly if it needs user input as it means there will be intra- and inter- observer variability. This means that the segmentation will not be reproducible as the threshold will change depending on the background region used. Threshold based segmentation methods are summarised in Table 1.5.

Study	Segmentation Type	Segmentation Methodology
Avril *et al.* (1997)	Manual Delineation	Manual segmentation on 2-D slices by an experienced clinician
Paulino and Johnstone (2004)	Fixed Threshold based on SUV	Fixed 2.5 SUV threshold
Erdi *et al.* (1997)	Threshold using % of SUV_{max}	3-D isocontour of ~40% of SUV_{max}
Boellaard *et al.* (2004)		3-D isocontour of 50% of SUV_{max}
		3-D isocontour of 70% of SUV_{max}
	Threshold using SUV_{max} and BG	3-D isocontour with threshold halfway (50%) between BG and SUV_{max} i.e. 0.5(Max + BG)
		3-D isocontour with threshold 70% of BG and SUV_{max} i.e. 0.7(Max + BG)
Black *et al.* (2004)	Threshold using SUV_{mean}	Threshold = 0.307 * SUV_{mean} + 0.588. Based on linear regression function from phantom experiments
Green *et al.* (2008)	Threshold using SUV_{mean}	Adaptive region growing method which uses a % of the SUV_{mean}, which changes with each iteration
Zasadny *et al.* (1998)	Threshold using BG	Threshold = mean of normal liver + 3 S.D
Wahl *et al.* (2009)	Threshold using BG	Threshold = mean of normal liver + 2 S.D., based on a VOI of 3cm in diameter in the right hepatic lobe of the liver
Gulec *et al.* (2011)	Threshold using BG	Threshold = SUV_{max} in the liver
Davis *et al.* (2006)	Threshold using SUV_{max} and BG	Threshold = BG + Relative Threshold (SUV_{max} – BG), relative threshold based on phantom measurements
Drever *et al.* (2006)	Threshold using SUV_{max} and BG	Threshold = Contrast Level (%) * (SUV_{max} – BG) + BG, based on phantom measurements
Nestle *et al.* (2005)	Threshold using SUV_{mean} and BG.	Threshold = (0.15 * SUV_{mean}) + BG where SUV_{mean} is taken from 70% of SUV_{max} segmentation
Nestle *et al.* (2007)	Threshold using SUV_{mean} and BG.	Threshold = (0.7 * SUV_{mean}) + (0.5 * BG) where SUV_{mean} is taken from 70% of SUV_{max} segmentation
Boucek *et al.* (2008)	Threshold using SUV_{mean} and BG.	Adaptive region growing method which uses a threshold called maximum normal level (MNL) = BG + 3 S.D. and SUV_{mean} to create a VOI of the tumour
Jentzen *et al.* (2007)	Iterative Thresholding	Iteratively calculates the optimal threshold based on signal-to-background threshold-volume curves derived from phantom measurements
Van Dalen *et al.* (2007)	Iterative Thresholding	Iteratively calculates threshold using relative threshold level (RTL) = (Threshold – BG) / (SUV_{max} – BG) where RTL is derived from phantom measurements
Nehmeh *et al.* (2009)	Iterative Thresholding	Iteratively calculates the optimal threshold based on target-to-background ratio and lesion volume using a Monte Carlo based mathematical model

Table 1.5: Comparison of Threshold Based PET Segmentation Methods

All segmentations use a threshold value to segment images. BG = Background, S.D. = Standard Deviation.

There are more sophisticated methods of PET tumour segmentation than thresholding methods. Traditional region growing techniques have been combined with dual active contours and have shown success at delineating tumours in PET phantom studies (El Naqa *et al*., 2007; Li *et al*., 2008). Gradient-based segmentation attempts to identify changes in the gradient in an image to detect borders of tumours. A gradient-based algorithm, applying the watershed transform and cluster analysis, has shown initial success on phantom studies (Geets *et al*., 2007). Probability based methods have been found to be accurate in segmenting PET tumours using a fuzzy locally adaptive Bayesian (FLAB) approach, showing good initial results and continued success on more data (Hatt *et al*., 2009; Hatt *et al*., 2010). Recent studies have investigated the use of artificial intelligence, in the form of neural networks, and possibility theory in delineating tumours on PET (Sharif *et al*., 2010; Dewalle-Vignion *et al*., 2011). While all these methods may provide a better solution to the problem of segmenting PET tumours in time, their development is at the early stages and they are difficult to implement into clinically based software. Table 1.6 summarises these more advanced segmentation methods.

Study	Segmentation Type	Segmentation Methodology
Geets *et al*. (2007)	Gradient Based Method	Gradient based method which uses the watershed transform and hierarchical cluster analysis
El Naqa *et al*. (2007)	Dual Active Contours	Uses dual active contours to identify edges of tumour and background to compute tumour delineation
Li *et al*. (2008)	Thresholding + Dual Active Contours	Uses adaptive thresholding based on volume changes at SUV_{max} %s and then applies dual active contours
Hatt *et al*., (2009)	Probability Theory Based Method	Uses probability theory in the form of a Fuzzy Locally Adaptive Bayesian (FLAB) based method
Sharif *et al*. (2010)	AI/Neural Network Based Method	Artificial neural networks are used in the wavelet domain to segment PET lesions
Dewalle-Vignion *et al*. (2011)	Possibility Theory Based Method	Uses possibility theory to avoid the use of binary values for edge voxels in a volume, using fuzzy c-means clustering to determine volume delineation

Table 1.6: Comparison of Non-Threshold Based PET Segmentation Methods

The issue of PET segmentation is made particularly problematic due to the lack of a clearly identified gold standard to compare delineation methods to. Segmentations are compared to phantom images, simulated data, segmented CT volumes and histopathology but all of these come with their own problems. Testing algorithms against spheres and phantoms, with well-defined homogenous uptake, may provide accurate results but cannot produce more realistic, heterogeneous lesions which are likely to occur in clinical data. While comparing PET segmentation to CT segmentation has the advantage of being applicable to clinical data, there is no way of knowing the accuracy of the CT segmentation itself or if it accurately represents the functional PET volume.

Histopathology measurements are the most accurate method for obtaining reliable TV measurements. However, histopathology does not give a perspective of whether the segmentation is in the correct position or shape in an image as well as often being impractical to obtain and in small samples when available. Simulated data, often using real clinical data as a starting point, is a favourable and practical way of testing segmentations. It can at least try to represent clinical data and there is a definite true TV before blurring and noise is added to the image, however, it is still only a simulation. Defining the 'best' segmentation method is always going to be flawed. Equally, the best segmentation method for one type of tumour, or disease, may not be the same for another. Segmenting smaller, homogeneous tumours in some types of lymphoma is going to be a very different task to segmenting larger areas of more heterogeneous disease in cancers such as mesothelioma. It may mean that different segmentation methods are needed for each. It has also not been fully investigated to what extent the segmentation method affects response measures.

1.3.5) Intensity Volume Histograms (IVH)

Image histograms are often used in image analysis and show intensity distribution over an image. Studies have used histograms to try and predict response using just pre-therapy PET scans and the comparison between pre- and post- therapy PET scans (El Naqa *et al*., 2009a; El Naqa *et al*., 2009b; Tixier *et al*., 2011; van Velden *et al.*, 2011; Vaidya *et al*., 2012). The first study to use parameters from histograms to assess response used a generalisation of the dose-volume histogram, named the intensity volume histogram (IVH) (El Naqa *et al*., 2009a). The IVH plots SUV against fractional volume of the VOI analysed, typically a tumour in the patient, and determines what volume of the tumour is over a given SUV. It effectively summarises the intensity information into a single curve for a particular area of disease (Figure 1.7).

Parameters of interest can be extracted from the IVH in an attempt to predict response such as percentage of TV above a given intensity. These parameters can be of potential use as they show intensity data which could be beneficial to the prediction of response which is not apparent using SUV_{max} alone. For example, a tumour with a high SUV_{max} and high SUVs in surrounding voxels may have a poorer response than a tumour with a high SUV_{max} with lower SUVs in surrounding voxels, as the latter case could represent a less aggressive tumour. While the SUV_{max} will remain the same in both these scenarios, parameters from a IVH would show this difference. IVH statistics have had promising initial results (El Naqa *et al*., 2009a; El Naqa *et al*., 2009b), but due to low patient numbers it is difficult to draw any real conclusions from the work. It has led to the use of IVH measures in other PET studies although their use in predicting response is still under investigation (Tixier *et al*., 2011; van Velden *et al.*, 2011; Vaidya *et al*., 2012).

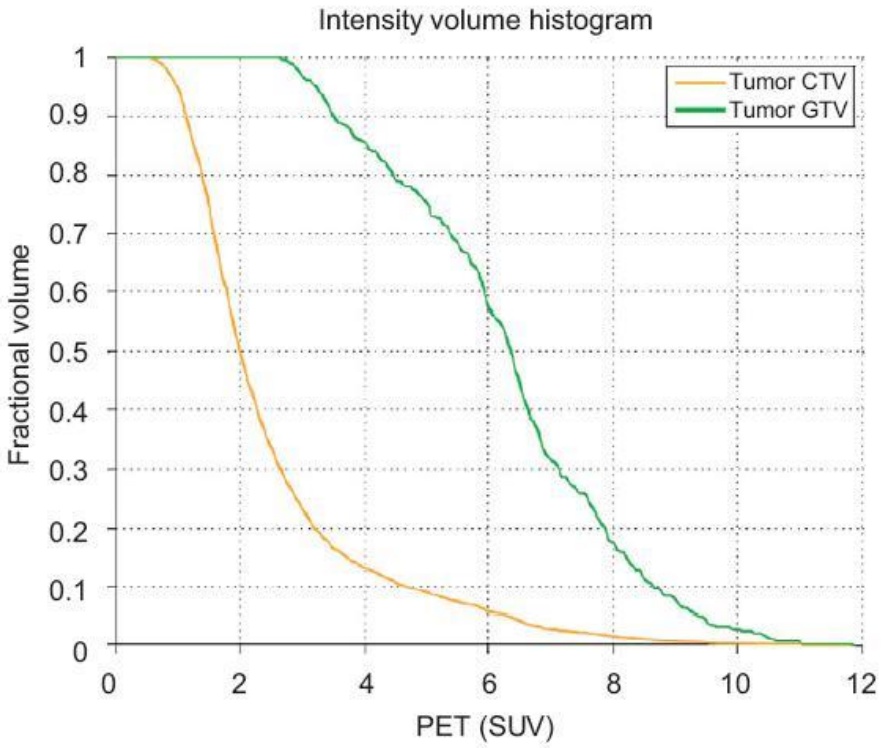


Figure 1.7: Example of an Intensity Volume Histogram (IVH)

In this example, the clinical target volume (CTV) (manually delineated by a physician) and the GTV (delineated using a 40% of SUV_{max} threshold, established for radiotherapy planning in a patient), are plotted (Taken from El Naqa *et al*., 2009a).

1.3.6) Texture Analysis

Texture analysis is a growing area in identifying response to therapy in PET studies and analyses texture properties of a tumour to see if changes can predict patient outcome (El Naqa *et al*., 2009a; El Naqa *et al*., 2009b; Tixier *et al*., 2011; Vaidya *et al*., 2012). This novel method of analysis highlights the pattern of uptake in a tumour, rather than intensity, and this can be highly specific and predictable in distinguishing between inflammatory uptake and disease (Hofman and Hicks, 2010). The most commonly used texture analysis method is the grey-level co-

occurrence matrix (GLCM), from which features such as contrast, energy, entropy, homogeneity and correlation can be extracted to assess response (Haralick *et al*., 1973).

The GLCM is a matrix defined as the distribution of co-occurring values at a given offset, originally designed to develop a set of features for classifying pictorial data (Haralick *et al*., 1973). Texture was defined to be the overall, or 'average', spatial relationship and distribution of grey tones over an image. If there is no pattern to an image, the variation of grey tones is wide and this results in a fine texture, whereas an image with more of a pattern has a coarser texture. Texture information of an image is specified by a set of grey-tone spatial-dependence matrices computed at various angular relationships and distances between neighbouring pixels. The co-occurrence matrix of a 2-D image can be defined as:

$$C\,\Delta x\,\Delta y\,(i,j) = \sum_{p=1}^{n}\sum_{q=1}^{m}\begin{cases} 1, \text{ if } I(p,q) = i \text{ and } I(p+\Delta x, q+\Delta y) = j \\ 0, \text{ otherwise} \end{cases} \qquad [1.8]$$

where C is the co-occurrence matrix, Δ_x and Δ_y are the offsets, n and m are the size of the image (i.e. the image is n x m pixels) and p and q are an image pixel in the n x m image. The values can be binary or a given type of greyscale (e.g. 8-bit, 16-bit or 32-bit colour).

The GLCM is constructed using four variables. The intensity values of the reference pixel (i) and neighbourhood pixel (j), the distance or offset of each neighbouring pixel (d) and the angle used (in °) from the reference pixel (a). For each angle and distance an i-by-j probability matrix can be obtained. The co-occurrence matrix is set out as a symmetrical grid of the number of grey levels in the image, so, for an image with 8 grey levels, it would be a 8-by-8 (i by j) matrix usually averaged for each of the four angles (Figure 1.8). Each distance would have a different matrix. Images often undergo quantisation to reduce the number of grey levels in them to make the GLCM more computationally efficient.

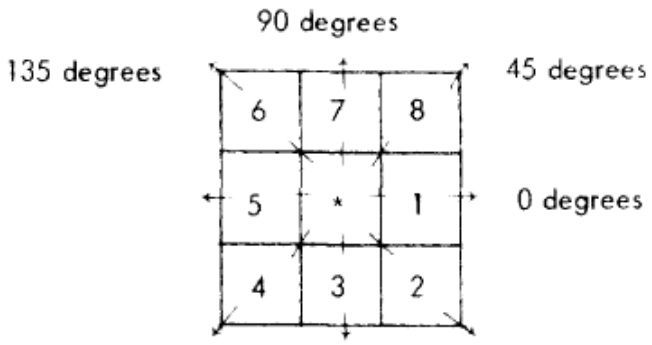


Figure 1.8: Possible Angles for Calculation of GLCM

In 2-D, a pixel has 8 neighbours opposite to it at different degrees. At 0° and 180° are pixels 1 and 5, at 45° and 225° are pixels 8 and 4, at 90° and 270° are pixels 7 and 3, and at 135° and 285° are pixels 6 and 2 (Taken from Haralick *et al*., 1973).

From these GLCMs, different parameters can be obtained which can represent texture features in the image. The most relevant are deemed to be energy, contrast, variance, correlation, entropy and the inverse difference moment/homogeneity (Baraldi and Parmiggiani, 1995). The GLCM was originally designed for 2-D images but with medical images predominantly being in 3-D a number of methodologies have been developed for computing the GLCM in 3-D images (Kurani *et al*., 2004; Tsai *et al*., 2007; Tesař *et al*., 2008). The GLCM has previously been used to assess texture in other medical imaging modalities, such as MRI, ultrasound and mammography (Garra *et al*., 1993; Lerski *et al*., Mudigonda *et al*., 2000; Mahmoud-Ghoneim *et al*., 2003; Alvarenga *et al*., 2007). More recently, it was used on the CT component of PET/CT scans as a potential marker for survival and correlated well with SUV_{max} and SUV_{mean} measurements in patients with oesophageal cancer and tumour survival in patients with NSCLC (Ganeshan *et al.*, 2012a, Ganeshan *et al.*, 2012b).

Other texture analysis measures useful for medical imaging include transform methods, which analyse the image in a different space, computationally exhaustive model-based methods, and other statistical approaches, such as gradient and run length matrices (Castellano *et al*., 2004). Of

these, grey level run lengths have been used to try and identify response in PET studies (Tixier *et al*., 2011). Grey level run lengths compute the number of times the same intensity is present along a run of pixels in an image at a specific angle resulting in a similar matrix to the GLCM from which parameters can be extracted (Galloway, 1975).

There are issues with using texture analysis to identify response in pre- and post- therapy PET scans as there are variations in results depending on image parameters, such as acquisition protocol and reconstruction parameters, which can affect most texture features by up to >30% (Galavis *et al*., 2010). However, a study comparing textural features between baseline PET scans just four days apart showed that some texture features, such as homogeneity and entropy, were very reproducible and that many texture features were as reproducible as SUV_{max} and SUV_{mean} (Tixier *et al*., 2012). Producing a reproducible methodology for texture analysis is difficult. Applying a fixed cubic VOI in the tumour will result in intra- and inter- observer variability, while using a semi-automated segmentation may eradicate this, an unsymmetrical VOI makes the computation of the 3-D GLCM more problematic. Despite these issues with reproducibility, there is a lot of interest in texture analysis in PET studies because of its success at distinguishing malignant and benign lesions (El Naqa *et al*., 2009a; El Naqa *et al*., 2009b; Tixier *et al*., 2011; Chicklore *et* al., 2012; Vaidya *et al*., 2012).

1.3.7) Shape Analysis

Just as texture analysis can distinguish between benign and malignant lesions so can shape (Adams *et al*., 1991; Rangayyan *et al*., 1997; O'Sullivan *et al*., 2003; O'Sullivan *et al*., 2005; Eary *et al*., 2008; El Naqa *et al*., 2009a). Studies of breast lesions on MRI and mammography images have shown more spherical and symmetric lesions are likely to be benign (Adams *et al*., 1991; Rangayyan *et al*., 1997). The success of shape analysis in these studies has led to its use in PET. Studies on sarcoma patients employed shape analysis using a measure based on the

difference between tumours from an idealistic elliptical sphere to characterise heterogeneity of tumours on PET images (O'Sullivan *et al*., 2003; O'Sullivan *et al*., 2005; Eary *et al*., 2008). It proved to accurately predict patient outcome in >200 patients. Different shape based features, including measurements for eccentricity, solidity and extent, were used to try and predict patient outcome on head and neck and cervix cancer patients (El Naqa *et al*., 2009a). Results in this study were variable with low patient numbers, but some shape analysis measures were more accurate than texture features. This is still an untested image analysis tool in identifying response to therapy, however, there is potential for it to be useful at predicting patient outcome.

1.3.8) Conclusion to Image Analysis in Response to Therapy

Image analysis has the potential to play a prominent role in analysing response to therapy in PET/CT studies. Unlike visual analysis, image analysis tools can provide quantitative and reproducible indices, which can be useful in predicting patient outcome, aiding clinician's reports and for standardising measurements in clinical trials. Research into using image analysis to extract quantitative features in images has been defined as 'radiomics' (Lambin *et al*., 2012). The growth of radiomics has led to the use of measures such as TV, TLG, IVH parameters, and texture and shape analysis on PET/CT studies as they may provide more information than SUV_{max} or SUV_{peak} values.

2) Software for Identifying Response to Therapy

2.1) Introduction to PETTRA Software

Identifying response to therapy between pre- and post- therapy PET/CT scans can be investigated using SUVs and volumetric measures, including SUV_{max}, SUV_{peak}, SUV_{mean}, TV and TLG. Additionally, more experimental methods such as IVHs, texture analysis and registered pre- and post- therapy subtraction images can also be used. Commercial software is available, including HERMES Volume Display and General Electric's (GE) Advantage Workstation, which can do some of these tasks. However, a tool to view and analyse PET scans specifically for analysing response would be advantageous. This led to the development of PET Therapy Response Assessor (PETTRA), a software tool designed using commercial software package MATLAB® 2012b (The MathWorks Inc., Natick, MA, 2000). PETTRA was custom built for analysing PET scans and extracting parameters for identifying response to therapy. This chapter assesses the development of PETTRA, the tools it offers and how it operates. PETTRA is capable of showing 3-D PET, CT and PET/CT images in all three planes (coronal, sagittal and transverse), segmenting PET images and calculating SUV, TV, TLG, and IVH parameters.

2.2) PETTRA Interface and Viewing Tools

2.2.1) PETTRA Interface

Software analysing any type of medical image needs to have a graphical user interface (GUI) to view the image in all planes, and the ability to navigate through them to see the full image. Rarely is analysing an image an automated process so it is important that an interface the user can interact with is accessible. PETTRA's GUI was created using the GUI Design Component (GUIDE) of MATLAB® software, allowing the designer to create an interface with text, labels, buttons, axes and other utensils to allow users to analyse PET images with ease (Figure 2.1).

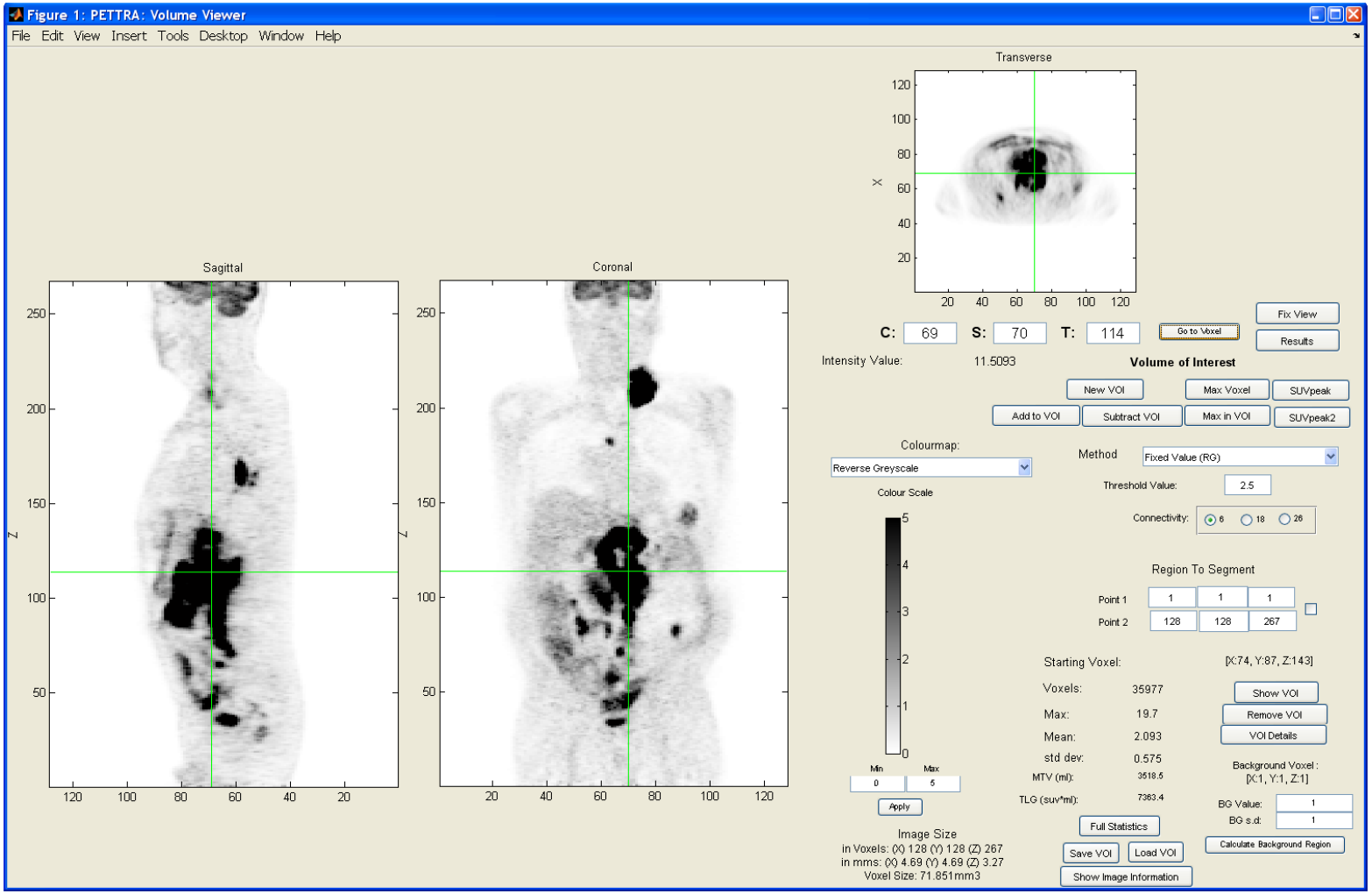


Figure 2.1: PETTRA User Interface

2.2.2) Viewing Images in PETTRA

Most medical imaging software tools allow the user to configure how they want to view and analyse images, allowing changes in image properties, and the way images are viewed. PETTRA displays data in the dimensions it is stored in, i.e. it does not interpolate data to fit the interface. The reason for this is that for assessing response to therapy, it is important that values are not affected by changes caused by interpolation. PETTRA has fixed axes in which data can be displayed, changing the height of the axis to keep the image in the right aspect ratio. PETTRA displays three axes, one for each of the coronal, sagittal and transverse planes. The coronal and transverse planes normally have the same image dimensions and are displayed together with the coronal plane in the centre of the display because this was the most viewed orientation during visual analysis for most image consultants. The transverse plane is typically square and so appears smaller on the viewer and is placed in the top corner to allow room for GUI components.

In PETTRA, users can click on any of the axes to select a specific voxel, highlighted by the green cursor lines on all three planes. The co-ordinates and intensity of the voxel are displayed. These co-ordinates can be changed by typing in new co-ordinates to highlight a voxel rather than clicking on the axes or using the arrow keys on the keyboard to go through the current axes slice by slice, giving the user a variety of ways to navigate through a 3-D image (Figure 2.1, Figure 2.2).

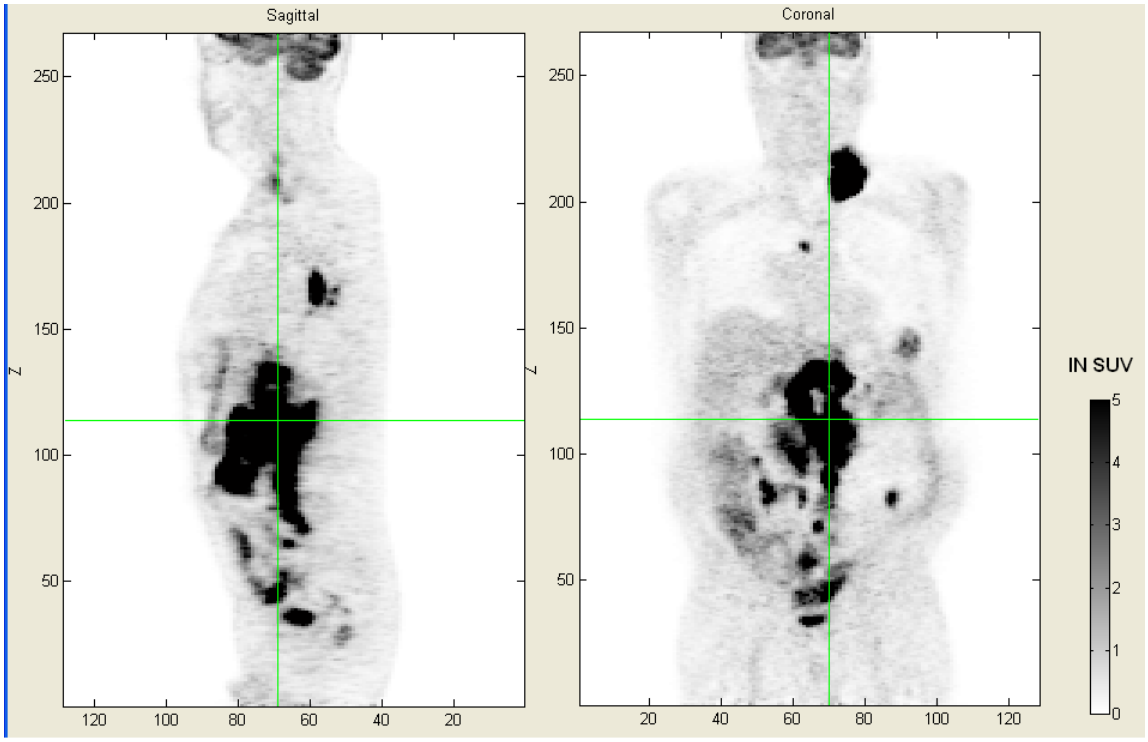


Figure 2.2: PETTRA 3-D Display and Navigation

The PETTRA GUI allows the user to navigate through the image by clicking the mouse on the axes, going through slices using the arrow keys on the keyboard. Green crosshiars show the position of the 3-D image in other planes.

Medical images are often viewed in a variety of different colour scales over different ranges of values, particularly when interpreting the intensity of tumours. Software should allow the user to customise the way the image is displayed. PETTRA allows the user to display the image in a number of different colour scales. The default is the reverse greyscale colour scale which is the preferred scale for most commercial imaging software, including HERMES. However, the user may choose to view the image in other colour scales including greyscale, hot, bone and jet

(Figure 2.3). Should a user want a specific colour scale, MATLAB® has functionality to create custom made colour scales which can be implemented in PETTRA.

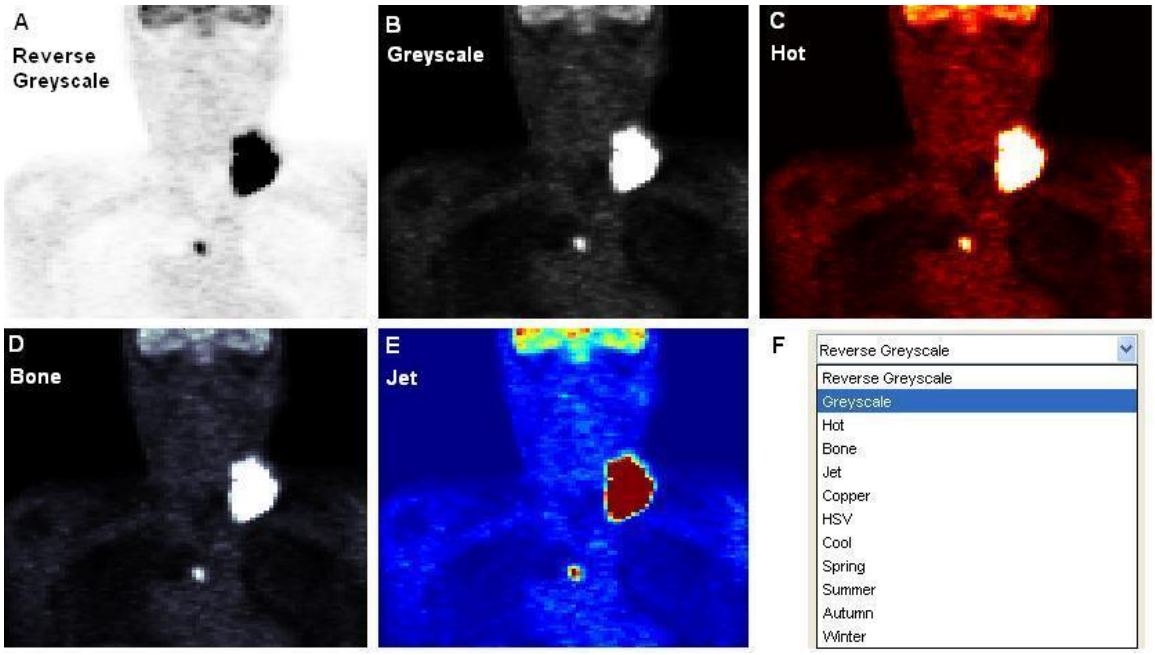


Figure 2.3: Colour Scale in PETTRA

Coronal view of a PET image displayed in (A) reverse greyscale, (B) greyscale, (C) hot, (D) bone, and (E) jet. The user can select the colour scale they want using a drop down box on the PETTRA interface (F). Colour range for all images is 0-5 SUV.

As well as colour scale, a user may also want to change the range of intensity in an image to aid image interpretation. By setting minimum and maximum values on a colour scale, all values over the maximum will be set to the specified maximum value and all values under the minimum will be set to the minimum value. This can be useful as it means that a user can, for example, modify background in an image by setting the minimum value to 1 SUV or set the maximum value to 2.5 SUV if it is believed that disease is >2.5 SUV. This allows the user to see

differences between background and disease in greater detail (Figure 2.4). The default setting for PETTRA is to use a minimum value of 0 and a maximum value of 5 on an image converted to SUVs.

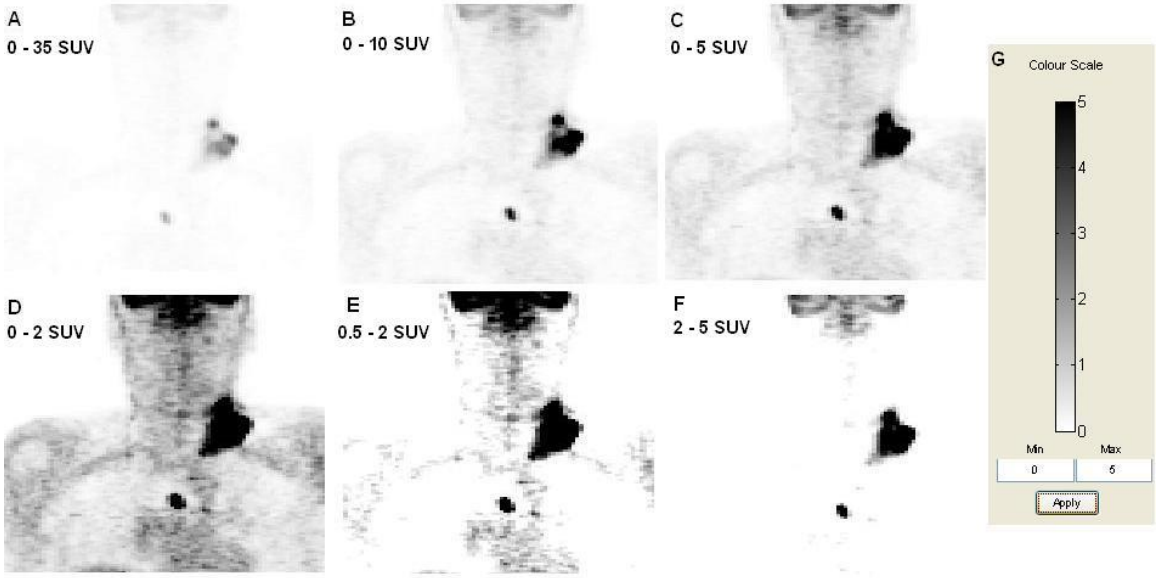


Figure 2.4: Colour Range in PETTRA

Coronal view of a PET image displayed in minimum to maximum SUV ranges of (A) 0 – 35.1 (SUV_{max} in the image), (B) 0 – 10, (C) 0 – 5, (D) 0 – 2, (E) 0.5 – 2, and (F) 1 – 5. The colour scale used is displayed on the PETTRA GUI with editable textboxes for changing the minimum and maximum values used (G). PETTRA's default display of images is in a reverse greyscale with a colour range of 0-5 SUV. PET images are displayed in this format in other figures unless otherwise stated.

2.2.3) Image Formats and Information

Medical images come in a variety of file formats from the universal Digital Imaging and Communications in Medicine (DICOM) standard to more specific formats designed for particular use at different institutions, such as the Guy's Image Processing Lab (GIPL) format. St Thomas PET Centre uses HERMES software which primarily stores images in Interfile 3.3 and DICOM. PETTRA is customised to deal with these file formats ahead of others. Interfile 3.3 is the preferred file format, although there has been development of software for loading both DICOM and Mayo Analyze formats too (Figure 2.5).

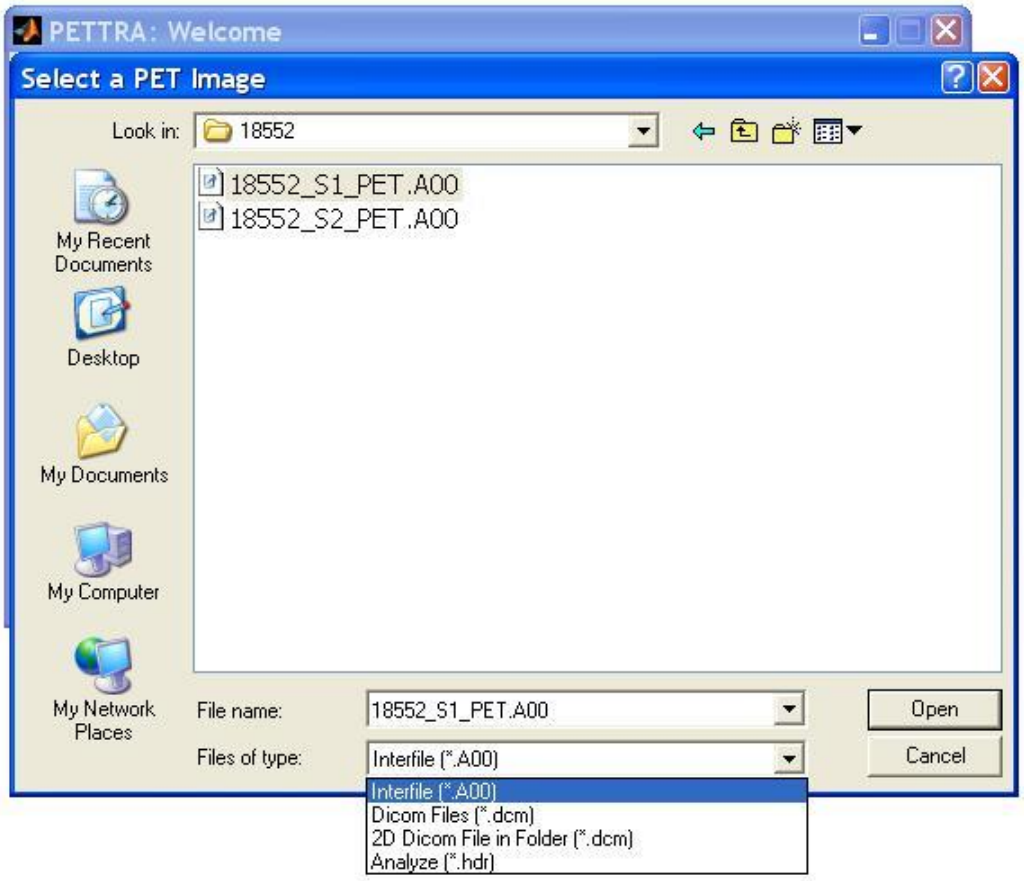


Figure 2.5: Loading Images in PETTRA

PETTRA can load images in Interfile 3.3, DICOM and Analyze file format selected by the user.

Analysis was performed on images in Interfile 3.3 format, however, as DICOM is the most common file format across many institutions and Analyze was used to register images with IRTK, these formats were considered when developing the software. It is possible to develop PETTRA to load in other formats if necessary. Along with the image, information in the image header file is also loaded into PETTRA as the information is vital in calculating SUVs and volumes. Header files contain information such as what institution the image is from, acquisition details, patient information and sizes/dimensions. This information can be important to the user and thus it is useful to display it if requested by the user. PETTRA can do this and the information shown can be easily adapted (Figure 2.6).

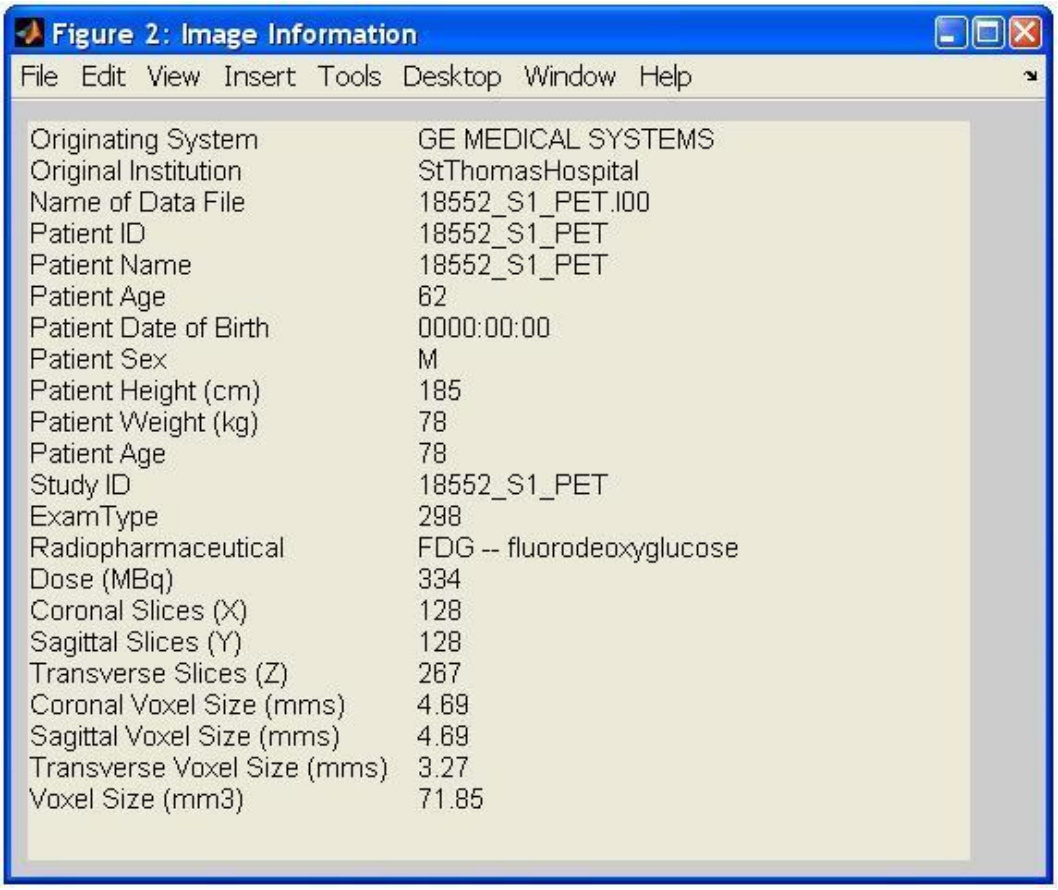


Figure 2.6: Image Information Table in PETTRA

PETTRA displays patient information, image measurements and other parameters if requested by the user.

2.2.4) Viewing Modes in PETTRA

PETTRA has several different modes for viewing images. The user can load up to two PET images and two CT images to analyse from the initial opening GUI. They can then choose to view images in a viewer for one image, two images, or two images and their resulting subtraction image (Figure 2.7). The subtraction image is most useful when two images are registered to allow voxel-by-voxel comparison. These viewing modes are designed with particular focus on the main aim of the tool, to assess response in pre- and post- therapy images, allowing two images to be viewed at once can aid this process. There is also a volume extractor tool for creating subvolumes of the image.

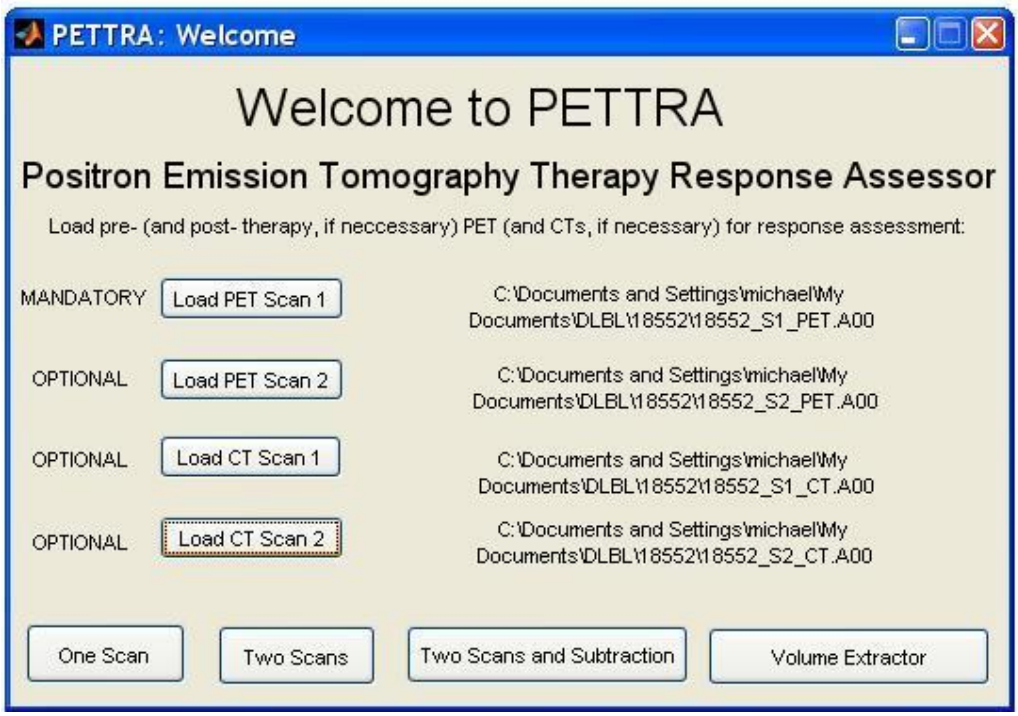


Figure 2.7: PETTRA Welcome GUI for Choosing Viewing Mode and Loading Images

User must load one PET image but can also load a second PET image and corresponding CT images. They can then view the image(s) in a viewer for just one scan, two scans, or two scans with a subtraction image. Additionally, the user can use a volume extractor for reducing image size for analysis or registration.

2.2.4.1) Viewing Pre- and Post- Therapy PET Images

The PETTRA two scan viewer allows pre- and post- therapy PET scans to be analysed side by side (Figure 2.8). The software is designed for two images with the same dimensions and voxel sizes to allow voxel-by-voxel comparison. Pre- and post- therapy PET/CT scans should ideally be acquired on the same scanner using the same protocol so the dimensions and, particularly, voxel size should be the same. There are cases where the dimensions are not the same and there are more slices in one image than the other, usually in the transverse plane. In this case, slices can be removed to allow images to be viewed. If the voxel sizes of the two images are different, interpolation can be done in MATLAB® so the images can be viewed in the software.

2.2.4.2) Viewing Pre- and Post- Therapy PET and Subtraction Images

The PETTRA two scan and subtraction image viewer shows both pre- and post- therapy images and their subtraction image (pre-therapy – post-therapy). The subtraction image can highlight areas in which a patient has responded, particularly if the post-therapy image has undergone registration. Parameters can be extracted from the subtraction image which could potentially be useful in identifying response. Figure 2.9 shows this viewing mode, where a patient has clearly responded between the pre-therapy and post-therapy image and this is visualised on the subtraction image where areas of black between liver and spleen show the change in disease. However, it should be noted that some of the changes are likely to be down to changes in physiological uptake and registration error. The default grey colour shows areas where no change between pre and post- therapy images has been observed.

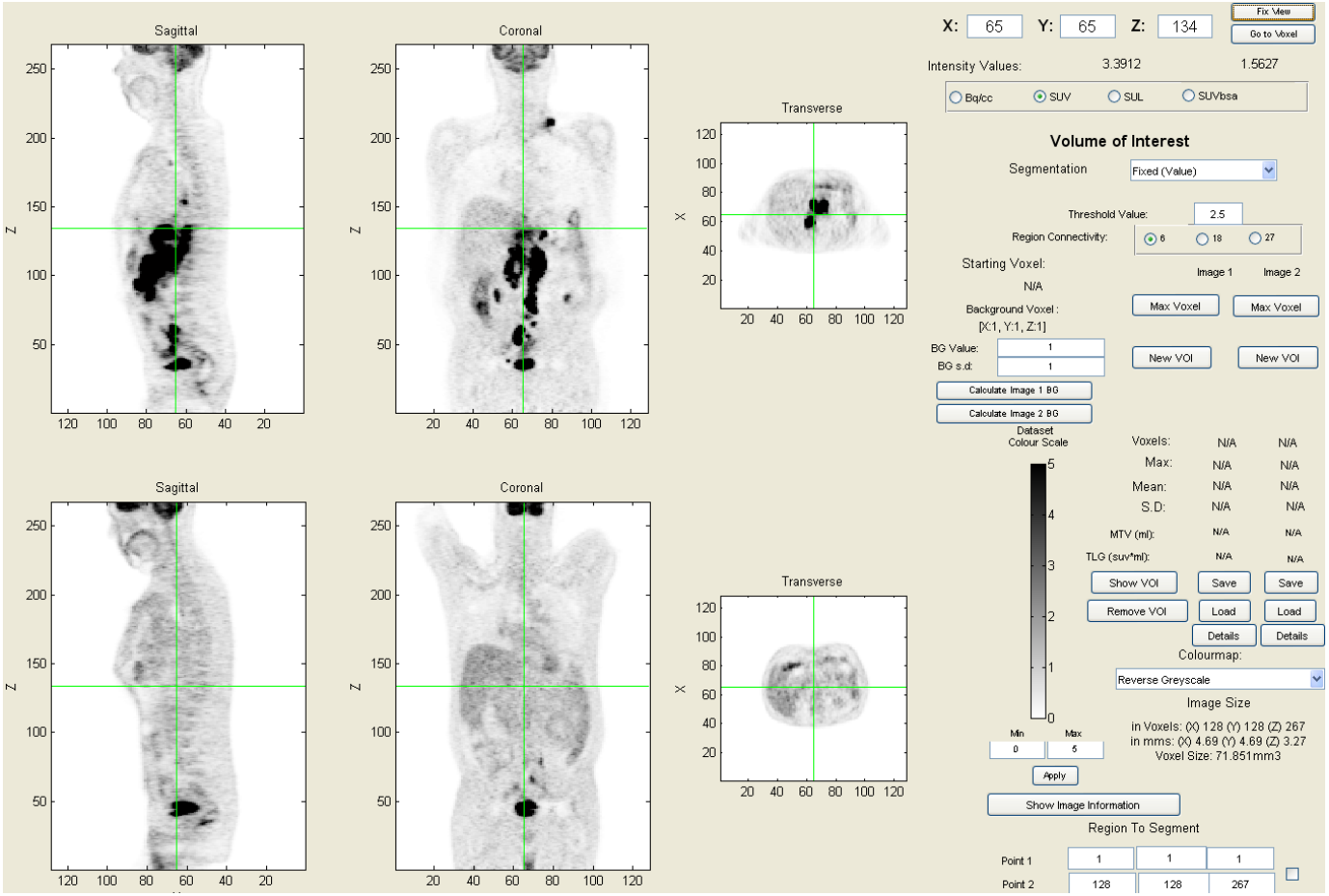


Figure 2.8: PETTRA Two Scan Viewing Mode

Two scan viewing mode, primarily designed to view pre- and post- therapy PET images together in one viewer. As with the one scan viewer, parameters for viewing can still be changed and images can be segmented and parameters for response extracted such as SUV_{max}, TV and IVH parameters.

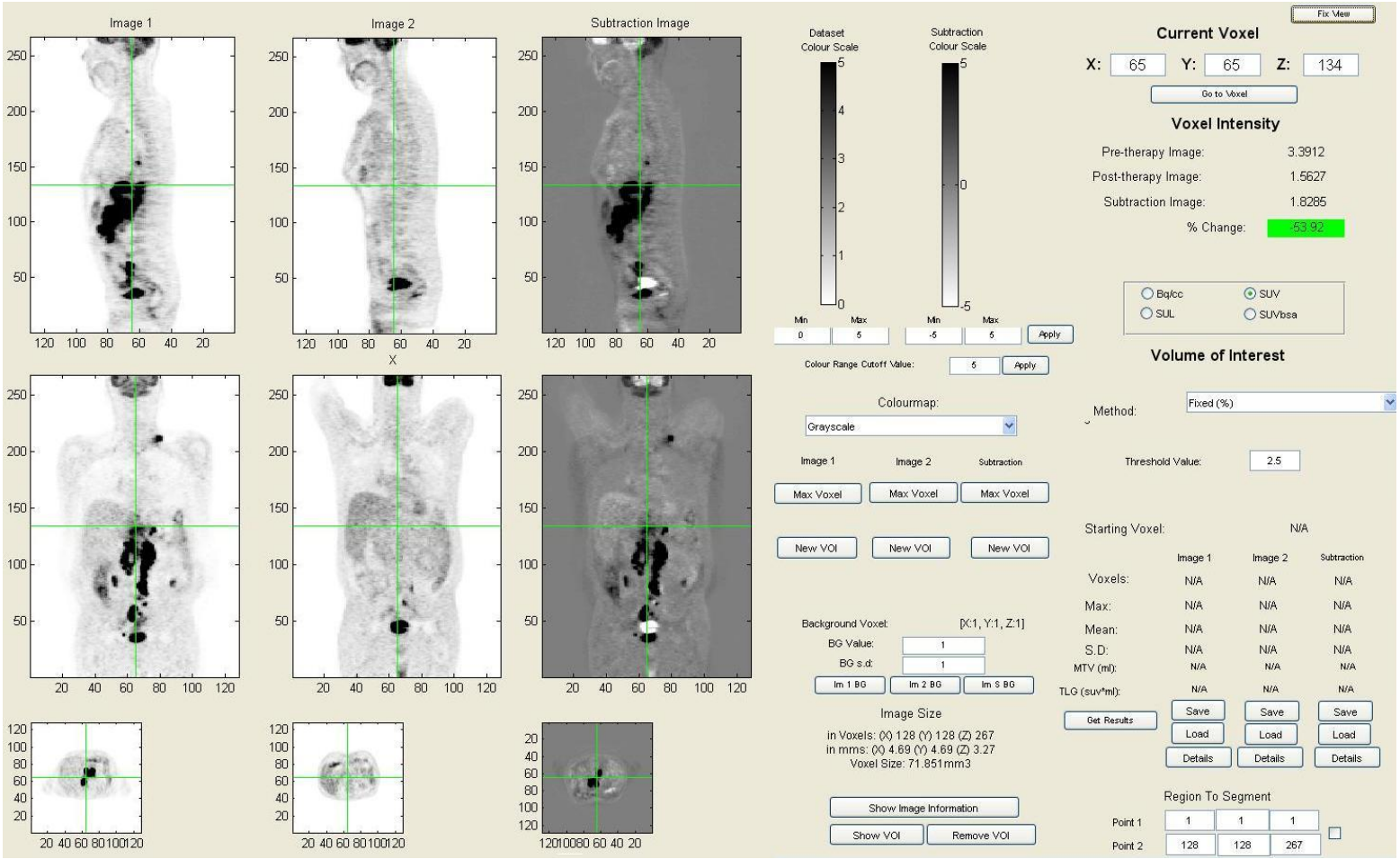


Figure 2.9: PETTRA Two Scan and Subtraction Image Viewing Mode

The two scan viewer with subtraction image is best used with registered images for voxel-by-voxel analysis. All the features of the one scan viewer are available.

2.2.4.3) Viewing PET and CT Images

While the viewer is primarily designed for analysing PET images, it can also display CT images, either individually or with corresponding PET images (Figure 2.10). CT images can be interpolated into the same image space as PET scans and overlaid in the viewer to allow the user to see where uptake on the PET scan corresponds, anatomically, on the CT scan. This is similar to the display in HERMES. For the analysis of data herein, PET images were viewed alone after areas of disease were segmented with the aid of a clinician using the more familiar HERMES software.

2.2.4.4) Subvolume Selection Tool

PETTRA allows the user to create subvolumes of pre- and post- therapy PET/CT images to focus on a specific VOI, such as a tumour, for analysing PET images or registering PET/CT datasets (Figure 2.11). The tool allows users to select two points in the images with the 3-D volume between the two points taken as the subvolume. The images can be saved in interfile format with the header file copied from the original images with the size of the image and new number of slices replaced to match any changes. These images can then be used for registration of subvolumes or analysis in PETTRA. Pre- and post- therapy PET scans are displayed together and the same subvolume is taken to ensure that registration can occur on images of the same dimensions and that a region can be taken which covers the main areas of interest on both images. The same approximate volumes can also be taken on the corresponding CT images.

Figure 2.10: CT Scan Viewed in PETTRA

In addition to PET images, CT images can be viewed in PETTRA in the same visual display.

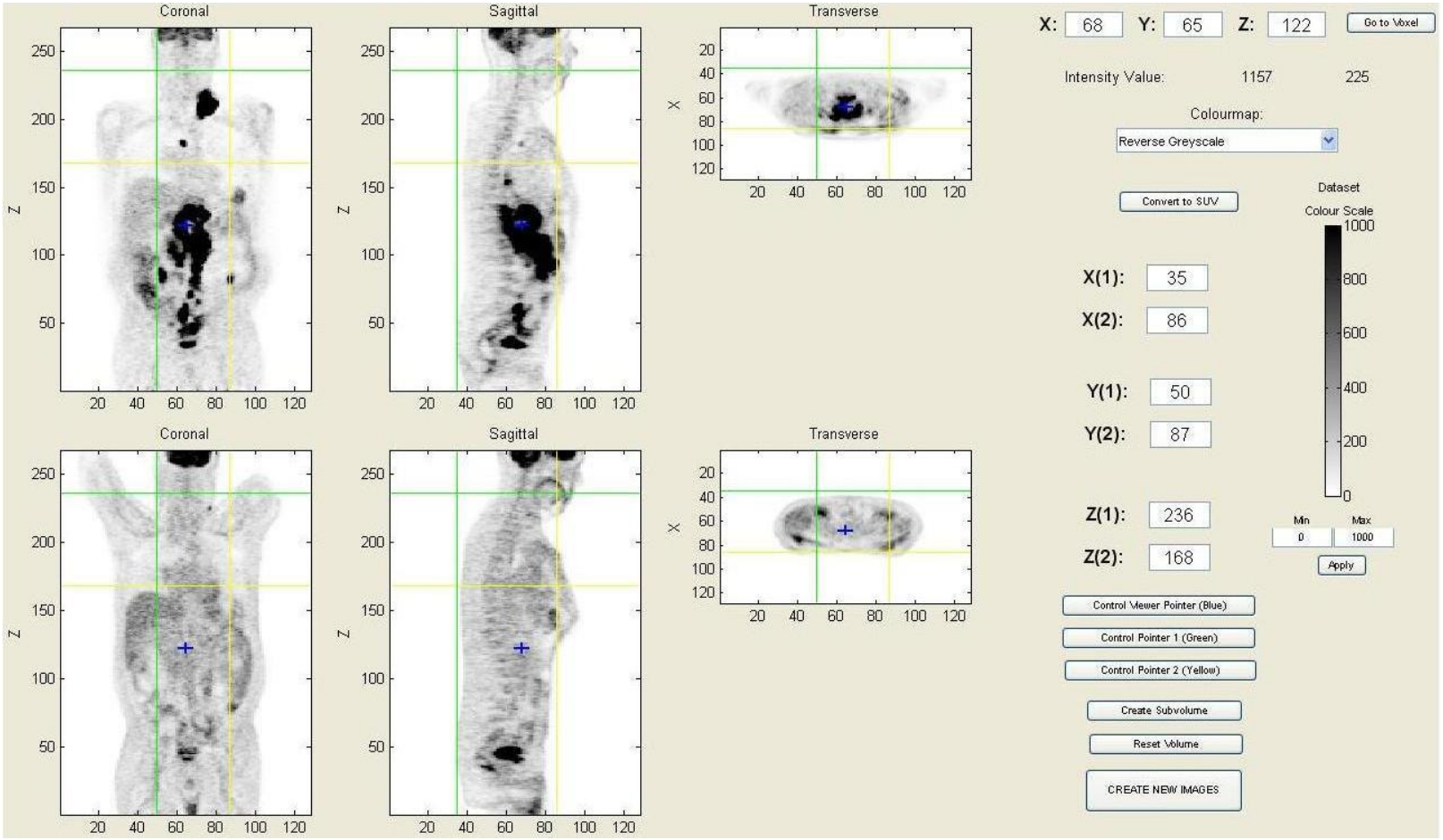


Figure 2.11: PETTRA Subvolume Selection Tool

The user can select two separate voxels in the images and the volume between these two points is taken as the subvolume, highlighted as the area encapsulated within the green and yellow lines on the images, both of which can be changed by the user. The blue crosshairs represent the voxel currently highlighted.

2.3) PETTRA Segmentation

2.3.1) Introduction to PETTRA Segmentation

With PET images displayed in PETTRA, one of the most important steps to allow assessment of response is the segmentation of areas of disease for analysis. PETTRA allows segmentation of disease using a number of methods, ranging from simple measures, such as placing a cubic volume over a region, to more complicated threshold methods which use SUV_{mean} of the tumour and a background region to adapt the threshold. The user can choose from a list of segmentation methods from a drop down box and change parameters in which the methods operate (Figure 2.12).

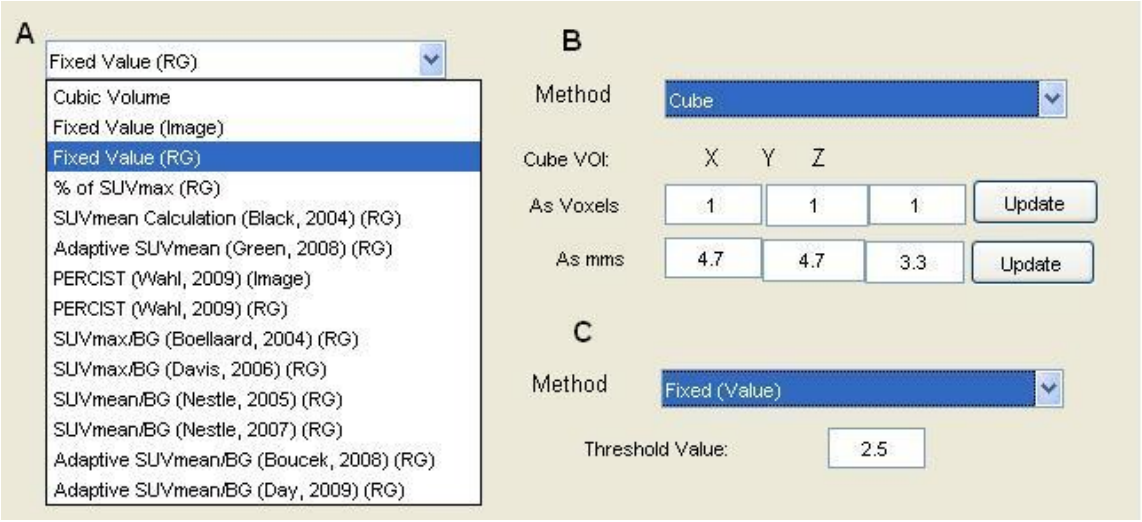


Figure 2.12: PETTRA Segmentation Methods

The user can choose from a list of segmentation methods (A) and adapt the parameters of them, such as dictating the size of a cubic region (B) and setting the threshold for a fixed threshold method (C).

The segmentation algorithms are all programmed to produce a binary mask of the image with the voxels included in the segmentation marked as true (value of 1) and those voxels not included as false (value of 0). An array of voxels in the segmentation is also included with the co-ordinates

in each plane recorded along with the intensity of the voxel, allowing the maximum, mean and standard deviation (S.D.) of the VOI to be calculated easily. Multiple segmentations can be performed on an image if there are multiple areas of disease with binary masks accumulated in a multidimensional array and voxels included added to the array list. PETTRA can combine the image with segmented VOIs to show the different areas of the image which have been segmented, with different colours representing multiple segmentations (Figure 2.13).

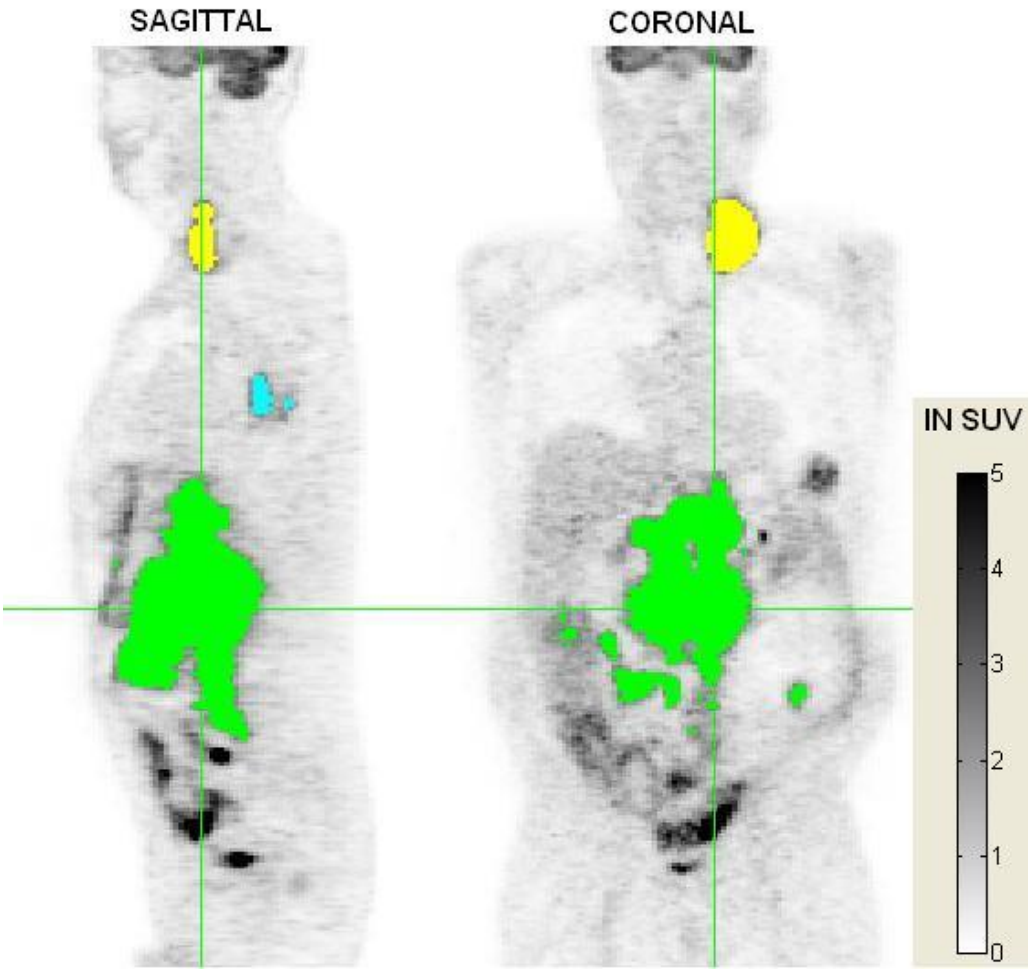


Figure 2.13: Visual Display of Segmentations of Disease on a PET Image

The PET image is combined with binary masks of segmented areas, represented by different colours to show different segmentations. The image with VOIs highlighted is shown in sagittal and coronal planes. As previously mentioned, images are displayed in a reverse greyscale colour scale and with a colour range of 0-5 SUV unless otherwise stated.

2.3.2) Segmentation of a Cubic Volume

One of the simplest segmentation methods implemented in the software creates a small cubic volume, typically used to segment a small area of tumour (Figure 2.14). This is useful in assessing distribution of voxels in an area of tumour using texture analysis, or similar measures, rather than for obtaining tumour volume (TV). The user can define the size of the cube based on how many voxels it uses from the initial, central voxel. For example, if three voxels are chosen in the coronal direction then the volume will include three voxels either side of the starting voxel, making a total of seven voxels. The user can select the displacement in all three planes using either voxels or mm. If mm are chosen rather than voxels, PETTRA will round up to the nearest mm which is a product of voxel size.

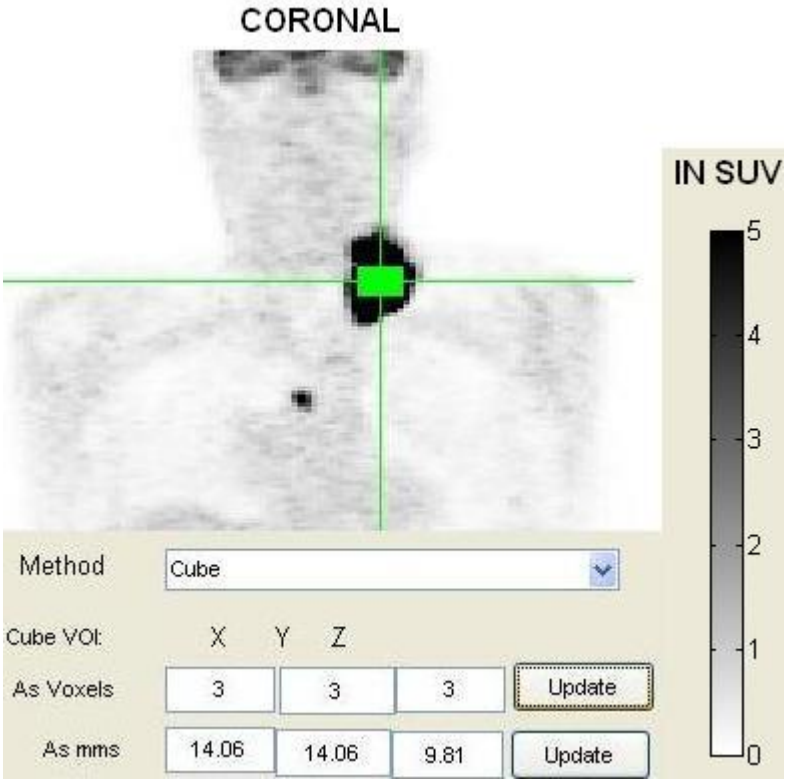


Figure 2.14: Visual Display of a Cubic Volume Segmentation

The user can choose the number of voxels included from the central, starting voxel in each plane in either voxels or mm. By pressing 'Update', PETTRA will display the values for the other units. If mm are used, they are rounded to the nearest multiplicative of voxel size.

2.3.3) Segmentation using a Fixed Threshold

Segmentation of PET images can be done using a fixed value of intensity to separate those voxels included in the VOI and those that are not. A fixed 2.5 SUV has been used to distinguish malignant and benign lesions (Paulino and Johnstone *et al*., 2004). PETTRA can create a VOI where all the voxels in an image over a certain fixed threshold value, editable by the user, are included and those under the threshold are not (Figure 2.15). The advantage of applying this threshold over the whole image is that it will pick up any area of disease, no matter how small, if it is over the threshold. The disadvantage is that unwanted areas of physiological uptake such as the brain, heart, liver, kidneys and bladder will be included and will need to be removed if the VOI is to be used for analysis.

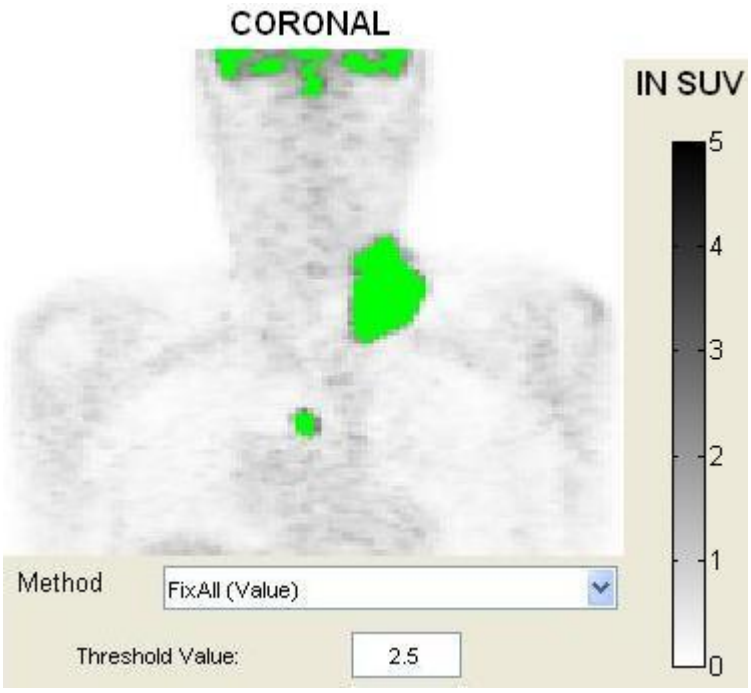


Figure 2.15: Visual Display of a Fixed 2.5 SUV Threshold Segmentation over an Image

In the coronal plane of the image, segmented areas show voxels >2.5 SUV. While most of this is disease, there is physiological uptake in the brain included.

Rather than perform a fixed threshold over the whole image, a region growing algorithm can be used to make sure that only voxels connected to a tumour are included in the VOI. Region growing starts from an initial voxel and then 'grows' a region over voxels connected to each other if they are over the threshold. This means a tumour can be segmented independently from the brain, unlike in Figure 2.15, unless the tumour and brain are connected by voxels over the threshold. Region growing continues to add voxels connected to those in the VOI which are above the threshold until no more can be added and the algorithm finishes.

Region growing is implemented in PETTRA and is used for several segmentation algorithms which use thresholding, including the fixed threshold method. Region growing starts with a seed point, selected as the current chosen voxel in PETTRA, determined by the user. This seed voxel is included in the VOI and its connected voxels are checked to see if they are over the given threshold. If they are, they are included in the VOI. If connected voxels are added to the VOI, their connected voxels are also checked to see if they are over the threshold and this process continues until all voxels have been checked. If no more connected voxels are above the threshold, the algorithm terminates (Figure 2.16).

The connected voxels to a voxel in the VOI can be defined in three different ways. Voxel connectivity can either use 6-connected voxels, 18-connected voxels or 26-connected voxels for a 3-D image (Figure 2.17). 6-connected voxels are those which touch the face of a voxel, 18-connected touch the face or edge of a voxel, and 26-connected voxels touch a face, edge or corner. PETTRA can perform region growing using 6-connected, 18-connected or 26-connected voxels, depending on the input of the user. The default is set to 6-connected voxels to avoid spillover into regions of physiological uptake.

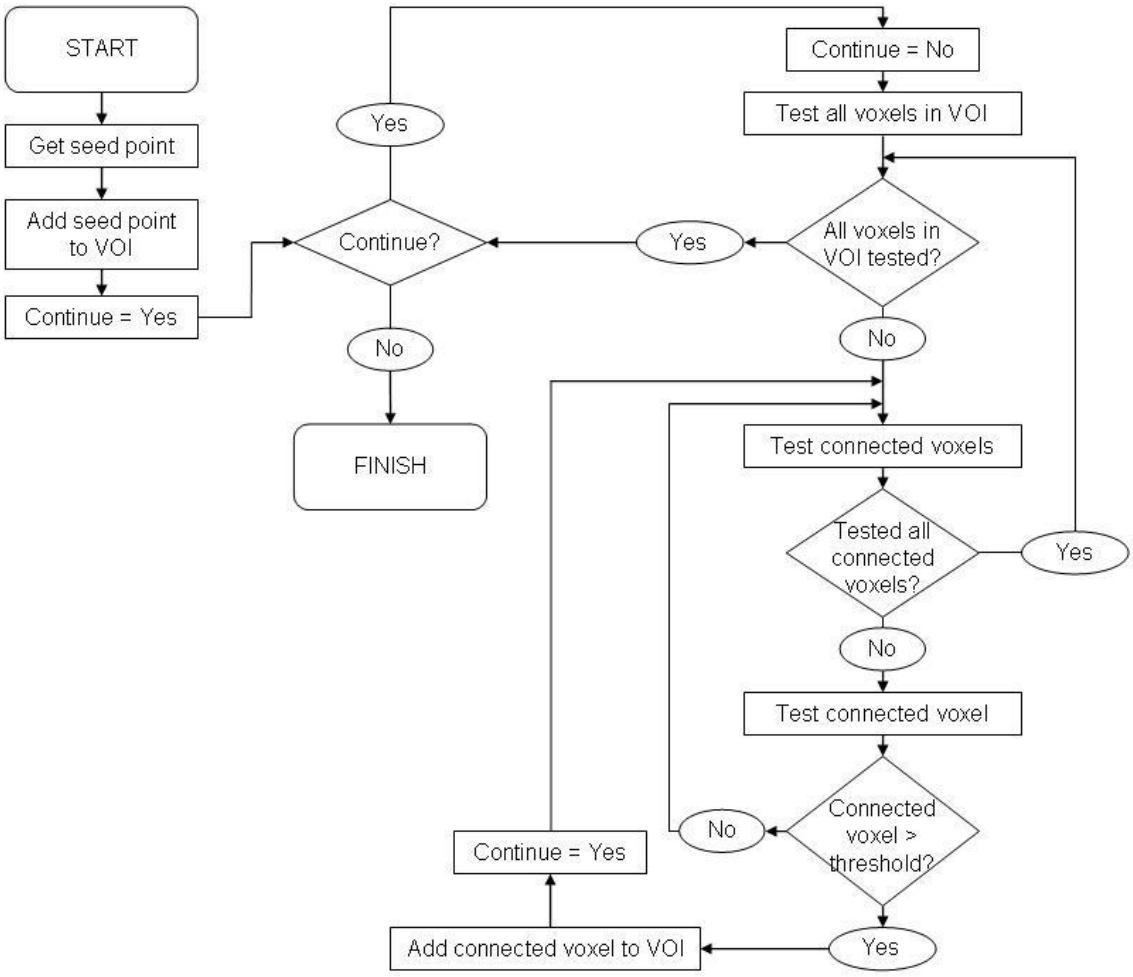


Figure 2.16: Flowchart for Region Growing in PETTRA

Region growing continues until the continue variable is left set to false after all the connected voxels have been checked to see if they are above the threshold. If no more connected voxels are above the threshold the algorithm will finish.

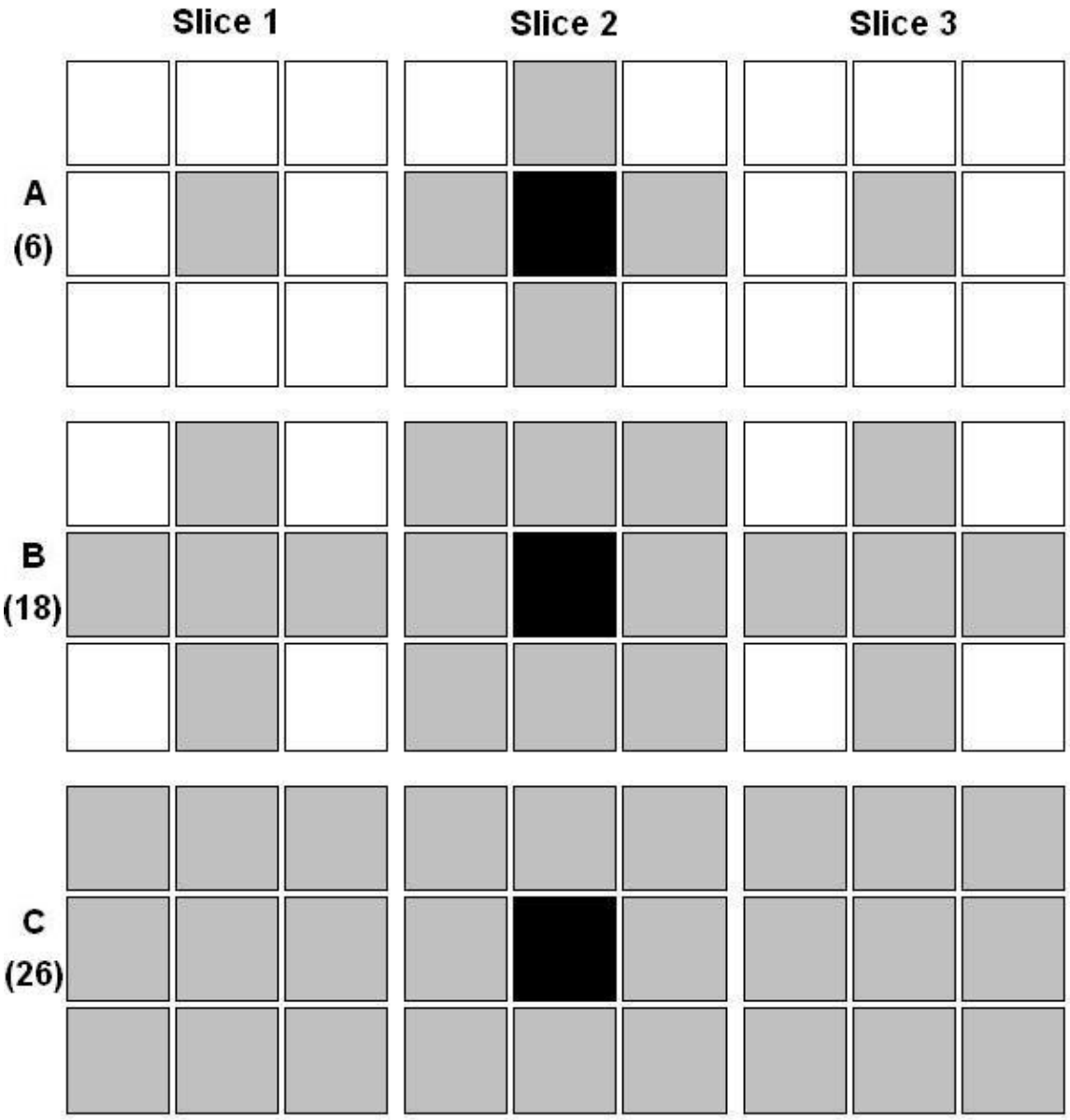


Figure 2.17: Voxel Connectivity for Region Growing Segmentation

The three types of voxel connectivity: (A) 6-connected, (B) 18-connected, and (C) 26-connected are illustrated over three slices of a 3-D image where the middle voxel in black is the one which voxels are connected to. Connected voxels are highlights in grey while non-connected voxels are white.

2.3.4) Segmentation using SUV_{max}

One of the most common segmentation methods in PET studies is using a percentage of SUV_{max} in a tumour as a threshold for delineation. The percentage used varies amongst PET studies with percentages from 40% to 70% used to segment PET tumours (Erdi *et al*., 1997; Boellaard *et al*., 2004). A region growing segmentation method using a percentage of SUV_{max} has been implemented in PETTRA, allowing the user to choose the percentage. To ensure the user does not have to choose the maximum voxel as the seed point, the tumour is first delineated using a region growing algorithm with a fixed 2.5 SUV threshold. When this VOI has been established, the maximum voxel is identified and threshold calculated as a percentage of it. The region growing algorithm is performed again using the new threshold. Figure 2.18 shows segmentations using the region growing algorithm with thresholds of 42% of SUV_{max} and 2.5 SUV.

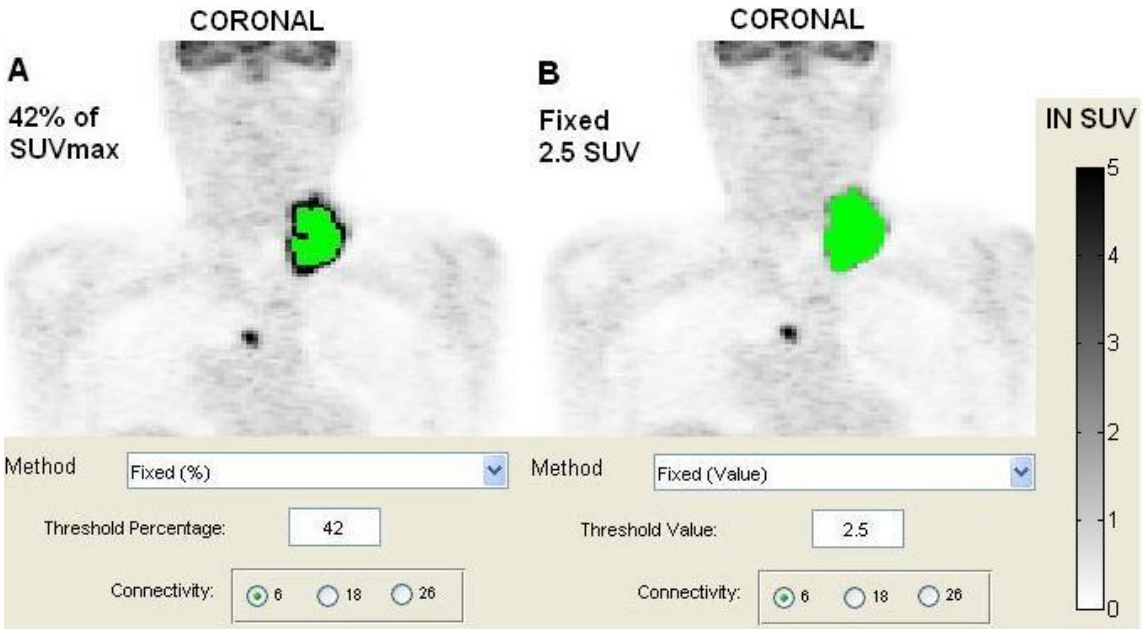


Figure 2.18: Visual Display of Region Growing Segmentations using Different Thresholds

The segmentations use thresholds of (A) 42% of SUV_{max} and (B) fixed 2.5 SUV. Both segmentations are of the same tumour and are done using region growing with 6-connected voxels.

2.3.5) Segmentation using SUV_{mean}

Rather than use a percentage of SUV_{max}, a linear relationship between the optimal SUV threshold and SUV_{mean} can be used to calculate a threshold based on phantom studies (Black *et al*., 2004). Using a linear regression function, the optimal SUV threshold was calculated as:

$$Threshold = 0.307 * SUV_{mean} + 0.588 \quad [2.1]$$

where SUV_{mean} is the mean uptake in the tumour and 0.307 and 0.588 are constant variables relating the SUV_{mean} to the threshold. A drawback of this method is the constant variables are likely to change depending on the scanner, and coming up with a SUV_{mean} for a TV of which volume is not know becomes a circular problem. However, this has been implemented in PETTRA by using a starting fixed threshold of 2.5 SUV and then continually modifying this threshold based on the calculation until it reaches a point where the change in TV is <1% or there have been at least 10 iterations.

An alternative segmentation method using SUV_{mean} uses a modified region growing method which continually adapts the threshold depending on the current SUV_{mean} of the VOI (Green *et al*., 2008). This region growing algorithm works differently to the standard region growing algorithm illustrated in Figure 2.16. Firstly, the algorithm is passed through twice, once to get an initial VOI to obtain the SUV_{max} and a second time with the SUV_{max} as the seed point for the algorithm. The threshold is recomputed each iteration as the SUV_{mean} from the current VOI is used to adapt the threshold throughout the algorithm (Figure 2.19). The percentage of SUV_{mean} used to calculate the threshold is user defined, but a percentage of 85% was found to segment tumours best in comparison with manual delineation by an experienced clinician. However, these were on images from a gamma camera for SPECT imaging so other percentages may be more appropriate for PET segmentation.

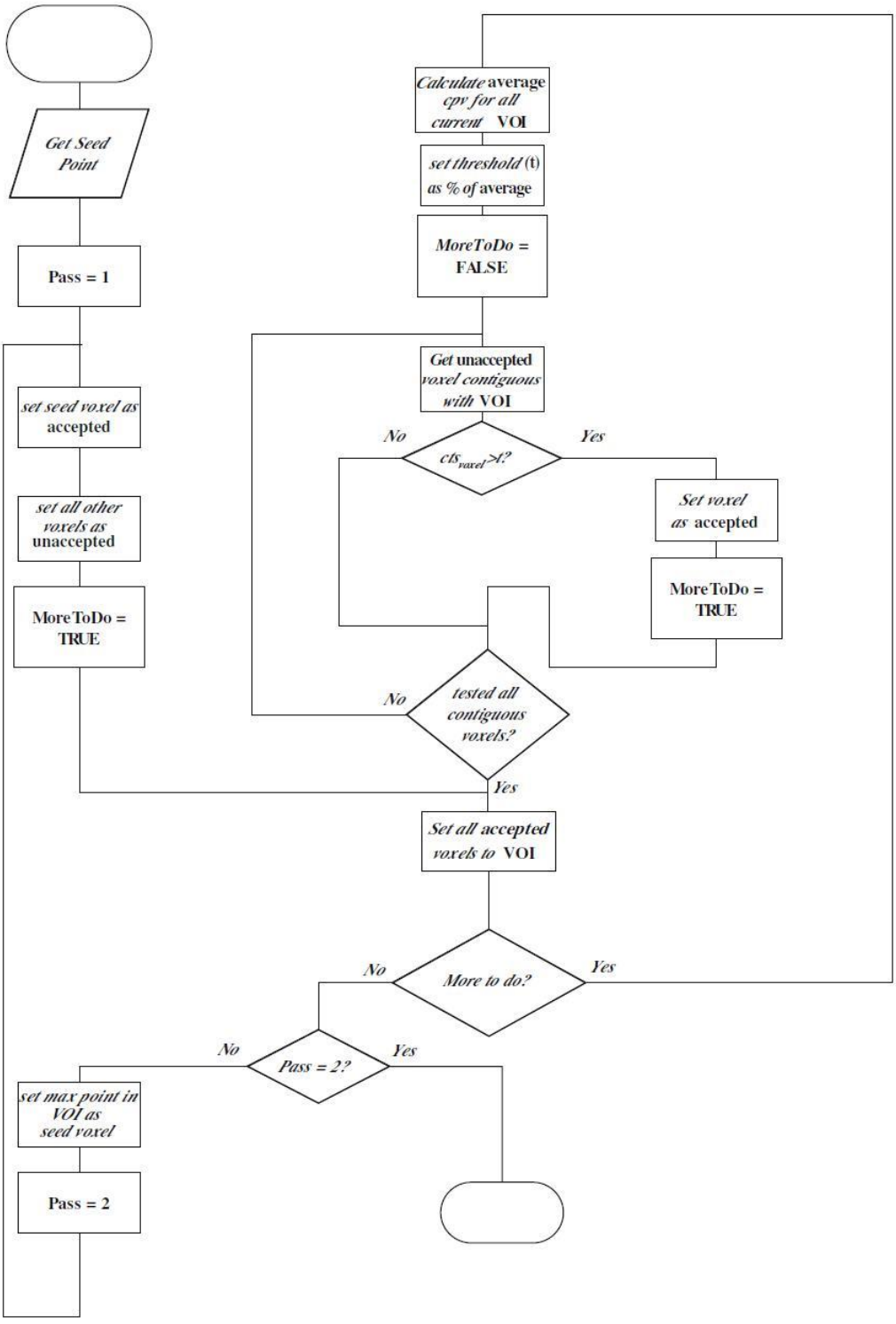


Figure 2.19: Flowchart for Adaptive SUV_{mean} Region Growing Algorithm

The algorithm computes the threshold during each iteration (Taken from Green *et al*., 2008).

Implemented in PETTRA, the region growing algorithm is adapted to match the flowchart in Figure 2.19 and the user can choose the percentage used to calculate the adaptive threshold. It is worth noting that other algorithms mentioned so far are all 100% reproducible, no matter where the user defines the starting point of the algorithm. The adaptive SUV_{mean} threshold can potentially yield different volumes depending on the starting point as the SUV_{mean} can be different depending on where the algorithm starts from. However, this is unlikely to be the case as the mean will no doubt end up being similar and segment a similar tumour volume with the same SUV_{max} which will then be used as the starting point for the second run through the algorithm. Both segmentation methods involving the SUV_{mean} are shown in Figure 2.20.

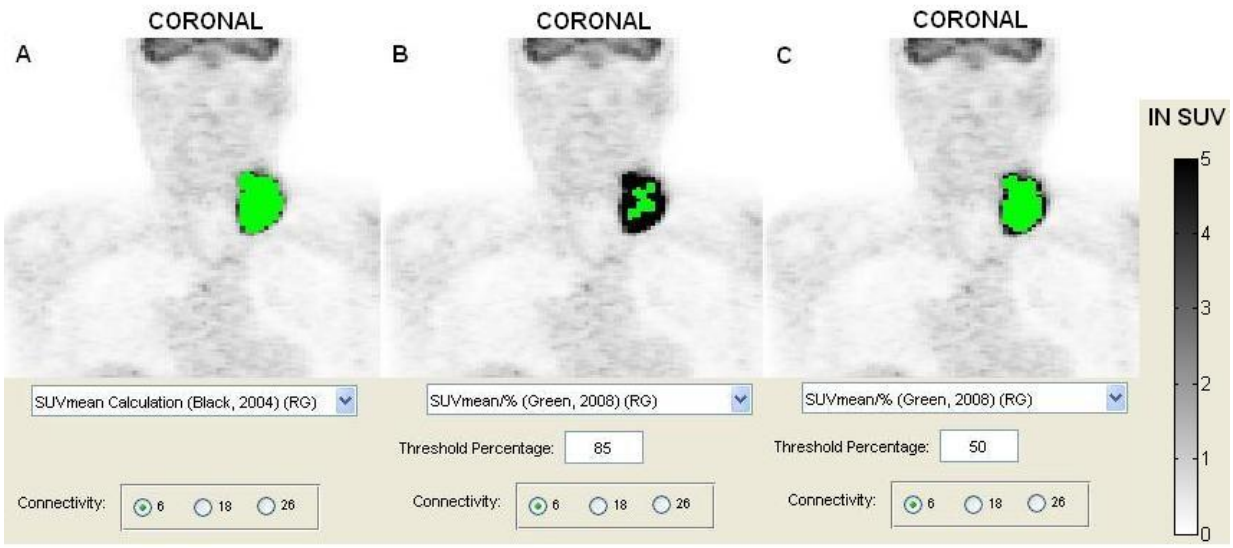


Figure 2.20: Visual Display of Region Growing Segmentations using SUV_{mean}

The segmentations use (A) Threshold = 0.307 * SUV_{mean} + 0.588 (Black *et al*., 2004), (B) Adaptive 85% SUV_{mean} (Green *et al*., 2008) and (C) Adaptive 50% SUV_{mean} (Green *et al*., 2008). It may be that a lower percentage threshold for the adaptive SUV_{mean} may yield more accurate segmentation for PET images, as in this image the 85% threshold only appears to segment the most intense part of the tumour (B).

2.3.6) Segmentation using Background Uptake

Some segmentation methods choose to use thresholds based on background activity, such as the method proposed in PERCIST criteria (Wahl *et al*., 1999). The guidelines state that for measuring TLG, TV should be obtained using a threshold based on the normal mean of the liver, calculated as:

$$\text{Threshold} = \text{Background Mean} + (\text{Background S.D.} * 2) \quad [2.2]$$

where Background is a spherical VOI with a 3cm diameter in the right hepatic lobe of the liver. For particularly active tumours, the S.D. may be multiplied by 3 rather than 2. The segmentation method is implemented in PETTRA using the formula in [2.2], and can be applied using region growing or over the whole image (Figure 2.21). The background region can be obtained using methods described in 2.3.10.

Other studies have used very similar segmentation methods based purely on background in an image (Zasadny *et al*., 1998; Gulec *et al*., 2011). PET segmentation using the PERCIST formula multiplied by 3 S.D. rather than 2 (Zasadny *et al*., 1998), and using SUV_{max} in the liver have both been proposed as thresholds for segmentation (Gulec *et al*., 2011). These methods can be applied in PETTRA as well as the PERCIST method. PETTRA allows the multiplication factor of the background S.D. in the PERCIST formula to be changed by the user and can use the fixed threshold method discussed in 2.3.3 to perform segmentation with a threshold equal to the SUV_{max} in the liver (Figure 2.21).

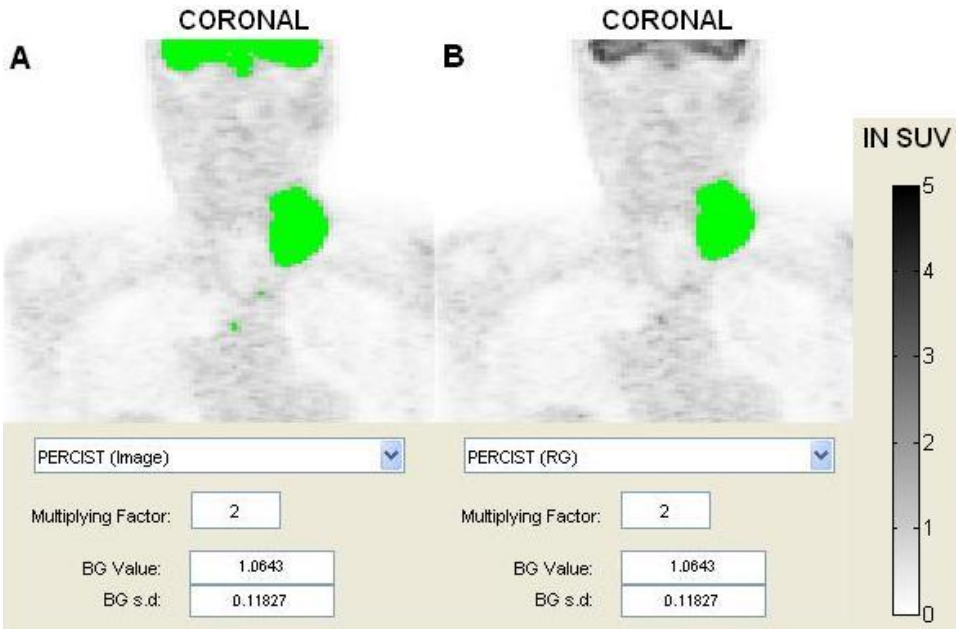


Figure 2.21: Visual Display of PERCIST Segmentations

PERCIST segmentation over the whole image (A) and on a single tumour using region growing (B) with background value/mean and S.D. for calculation chosen by the user. The multiplying factor, also editable by the user, defines multiplication of S.D. in the PERCIST formula so either method of calculating background + S.D. can be used (Zasadny *et al*., 1998; Wahl *et al*., 2009). Connectivity was set to the default 6-connected voxels for the region growing algorithm.

2.3.7) Segmentation using SUV_{max} and Background Uptake

Segmentation methods can use a background region along with SUV_{max} to calculate a threshold. A threshold based on 50% or 70% of the combination of SUV_{max} and a background region has been suggested (Boellaard *et al*., 2004), calculated as:

$$\text{Threshold} = 0.5 \text{ or } 0.7 \; (SUV_{max} + \text{Background}) \qquad [2.3]$$

where 'SUV_{max}' is the maximum voxel in the tumour and 'Background' is a mean value taken from a background region in the image. The advantage of this method is the higher the uptake in

the background, the higher the threshold will be for segmentation, with the aim to ensure that segmentation does not spill into areas of background around the tumour. This segmentation method is implemented in PETTRA using region growing and a fixed 2.5 SUV segmentation is used to obtain the SUV_{max} of the tumour. The percentage (50%, 70% or another percentage) and background mean are chosen by the user. Experimentation with different percentages may be useful on clinical data as the formula was devised on phantom experiments (Figure 2.22).

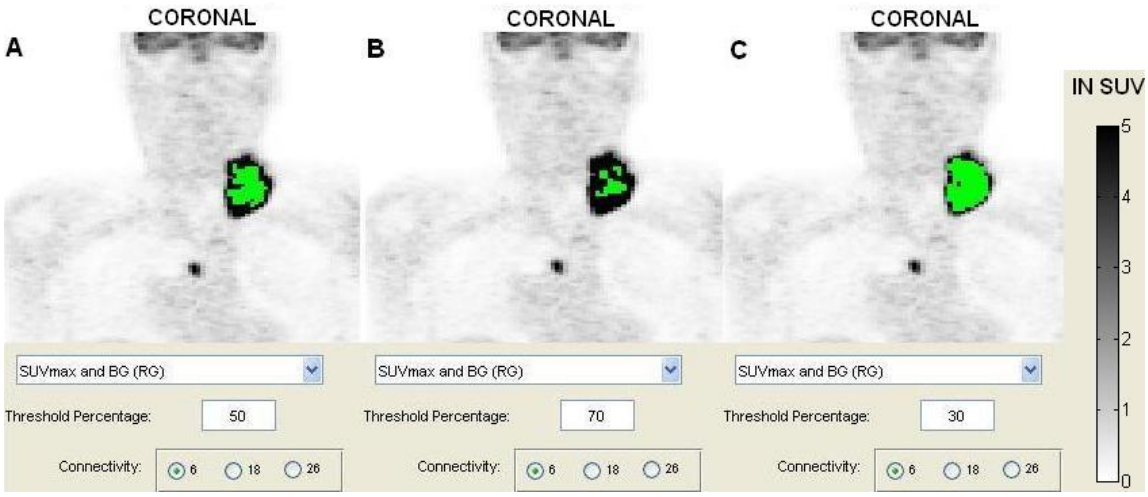


Figure 2.22: Visual Display of SUV_{max} and Background Segmentations

Segmentations are on the same tumour for (A) 50%, (B) 70% and (C) 30% of SUV_{max} + Background

Another segmentation method proposes using a background corrected SUV_{max} calculation to obtain a suitable threshold (Davis *et al*., 2006). The threshold is defined as:

$$\text{Threshold} = \text{Background} + \text{Relative Threshold} \left(SUV_{max} - \text{Background} \right) \quad [2.4]$$

where 'SUV_{max}' and 'Background' are as described before and the Relative Threshold is a fixed percentage of background subtracted SUV_{max}, dependent on the size of the tumour/object. In a phantom study, for spheres with diameters >12.5mm, a relative threshold of 0.41 (41%) was

found to be most accurate for segmenting volume (Davis *et al*., 2006). However, thresholds varied for those with diameters <12.5mm. For tumours where the diameter was >12.5mm, the SUV_{max} was taken as the mean of the highest 10% of adjacent pixels, similar to a SUV_{peak} measure.

Implementation of the algorithm in PETTRA is similar to the previously described SUV_{max} and background segmentation (Boellaard *et al*., 2004). The user chooses the background and relative threshold value while SUV_{max} is taken from a fixed 2.5 SUV segmentation from the initial seed voxel. In the study by Davis *et al*. (2006), an algorithm to determine the volume of the lesion is used to customise the relative threshold used according to phantom measurements (Davis *et al*., 2006). This has not been implemented in PETTRA, however, if phantom measurements are available, a tumour volume size can be obtained using an initial segmentation and the relative threshold changed to match these results. Once again, experimentation with relative thresholds may be useful on clinical data as the formula was devised primarily on phantoms (Figure 2.23).

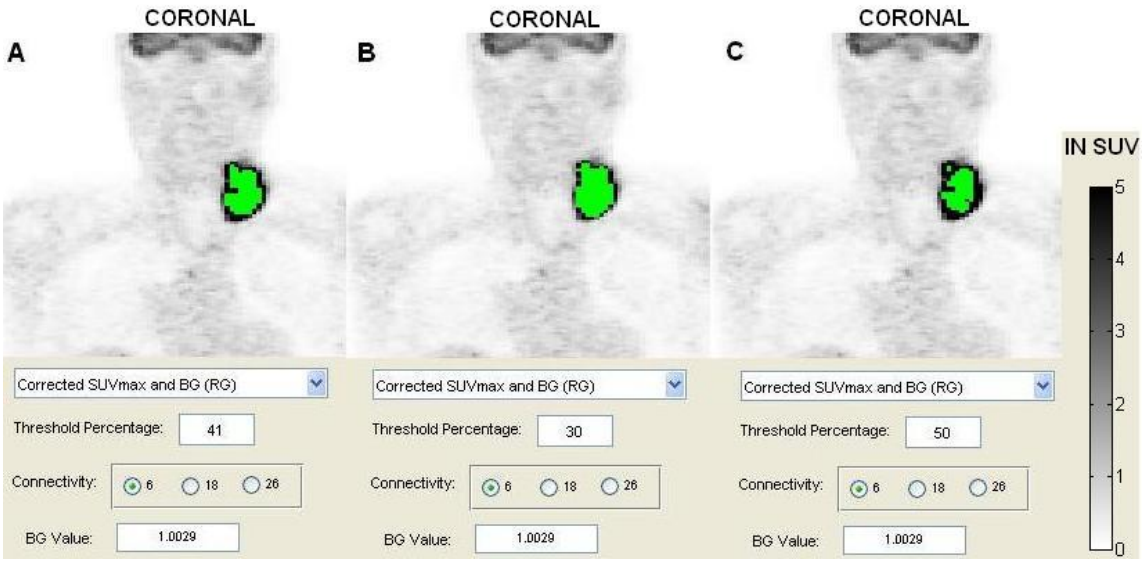


Figure 2.23: Visual Display of Background Subtracted SUV_{max} Segmentations

Relative thresholds of (A) 41%, (B) 30% and (C) 50% are used on the same tumour.

Other investigations have studied the relationship between SUV_{max}/SUV_{mean} and background with both tumour size and T/B ratio (Daisne *et al*., 2003; Yaremko *et al*., 2005), and these can be related to formula [2.4] (Davis *et al*., 2006). A study by Drever *et al*., (2006) uses the same formula but with the relative threshold changed depending on contrast rather than tumour size (Drever *et al*., 2006). Some studies have used iterative thresholding to segment PET images using similar formulas and phantom experiments, but using algorithms which continue to recalculate the threshold value and tumour size and/or contrast ratio until there is no significant change in values (Jentzen *et al*., 2007; Van Dalen *et al*., 2007; Nehmeh *et al*., 2009). While these segmentation methods could all be implemented in PETTRA, they are heavily related to the threshold calculation in [2.4] and all these methods use phantom data which is unlikely to be available in many cases. Therefore, addition of these sorts of methods in PETTRA seems redundant.

2.3.8) Segmentation using SUV_{mean} and Background Values

Rather than using SUV_{max}, segmentation based on tumour and background intensities can use SUV_{mean} of the tumour (Nestle *et al*., 2005; Nestle *et al*., 2007; Boucek *et al*., 2008). Nestle *et al*. (2005) defined a threshold for segmentation as:

$$\text{Threshold} = (0.15 * SUV_{mean}) + \text{Background} \qquad [2.5]$$

where SUV_{mean} is the mean of the tumour segmented using an isocontour of 70% of SUV_{max}, Background is the SUV_{mean} of a ROI placed over a relevant background structure, and 0.15 is a constant based on phantom measurements. The background VOI is advised to be placed over an anatomic entity, such as mediastinum or liver, adjacent to the tumour but a far enough distance away to ensure no disease is included in the VOI. The highest uptake of the background regions should be used to ensure the threshold is high enough to avoid spillover into other areas of physiological uptake, such as the liver, spleen, kidneys, bowel and bladders.

Once again using phantom measurements with further testing on clinical data, Nestle *et al*. (2007) revised the method for calculating the threshold through this technique (Nestle *et al*., 2007), with the new threshold being defined as:

$$\text{Threshold} = (0.7 * SUV_{mean}) + (0.5 * \text{Background}) \quad [2.6]$$

where SUV_{mean} is the mean of the tumour segmented using an isocontour of 70% of the SUV_{max}, Background is the SUV_{mean} of a ROI of 20 - 60cm^3 in the mediastinum and 0.5 and 0.7 are constant variables based on phantom measurements.

Another segmentation method, named the GRAB algorithm, is based on an iterative region growing algorithm using a percentage of SUV_{mean} (Boucek *et al*., 2008). GRAB uses both SUV_{mean} and background to calculate a threshold to use to segment the tumour. It uses a product of the SUV_{mean} in the tumour and a threshold factor, which uses the maximum normal level (MNL) to define the threshold. The MNL is based on the equation used for obtaining a threshold in the PERCIST segmentation [2.2], but uses a multiplying factor of 3 rather than 2 for S.D. The MNL is combined with the SUV_{mean} to provide the threshold factor.

$$\text{MNL} = \text{Background Mean} + (\text{Background S.D.} * 3) \quad [2.7]$$

$$\text{Threshold Factor} = 1 - \frac{(SUV_{mean} - \text{MNL})}{(SUV_{mean} + \text{MNL})} \quad [2.8]$$

$$\text{Threshold} = SUV_{mean} * \text{Threshold Factor} \quad [2.9]$$

Calculations in [2.5] and [2.6] are implemented in PETTRA in the same way as the method by Black *et al*. (2004), starting with a segmentation of a fixed 2.5 SUV threshold and iteratively

using the relevant calculation until it reaches a point where TV changes is <1% or there has been at least 10 iterations. The GRAB segmentation method uses the same coding in MATLAB® as the algorithm by Green *et al*. (2008) but changes the threshold calculation to use the threshold factor (Green *et al*., 2008), calculated by the MNL (Figure 2.24).

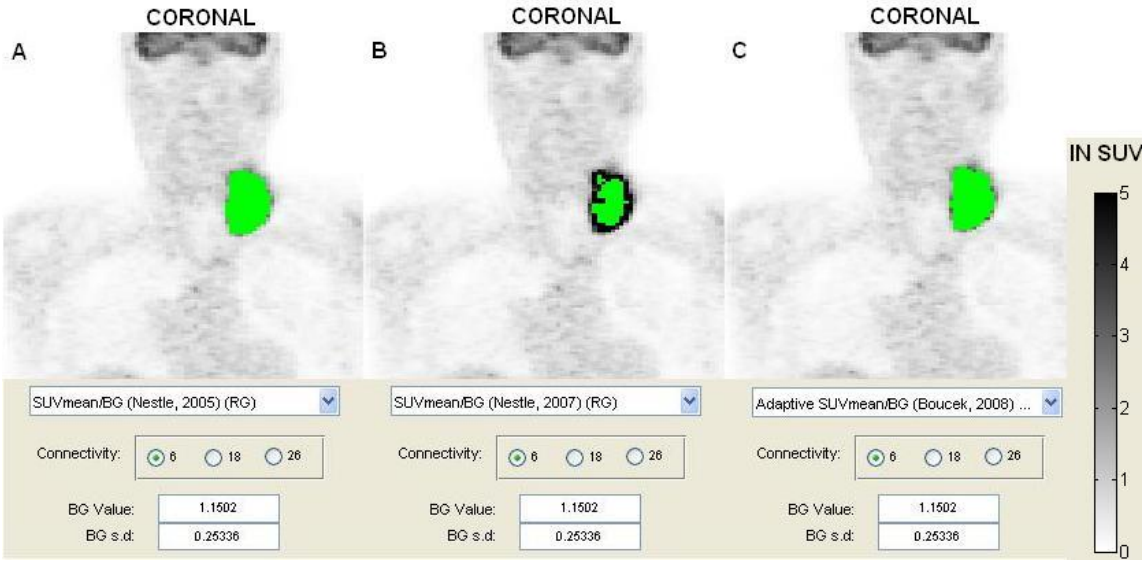


Figure 2.24: Visual Display of Segmentations using SUV_{mean} and Background

Segmentations using (A) Threshold = (0.15 * SUV_{mean}) + BG (Nestle *et al*., 2005), (B) Threshold = (0.7 * SUV_{mean}) + (0.5 * BG) (Nestle *et al*., 2007), and (C) GRAB algorithm (Boucek *et al*., 2008), on the same tumour. BG = Background.

2.3.9) Restricting Segmentation

On a PET image, there can be many areas of physiological uptake that can interfere with segmentations. Region growing algorithms are designed to not 'spill' over from segmenting tumour and include areas of physiological uptake but sometimes this is inevitable. For example, a segmentation of mesothelioma can include the heart through no fault of the segmentation method, as both the heart and disease can have areas of high uptake and are connected by

adjacent voxels (Figure 2.25A). Semi-automated segmentation methods are designed to have user input and therefore this will be noted when segmenting the image. However, the user needs a method to remove the area of physiological uptake or to avoid segmenting it initially.

PETTRA has two functionalities that allow the user to customise the segmented VOI. Firstly, the user can remove segmented regions as well as adding them. Removing a region takes all the voxels in the newly segmented region and subtracts them from any previously segmented VOIs and the list of voxels included in the VOI is updated. This allows the user to segment disease and then remove any physiological uptake. Secondly, PETTRA allows the restriction of segmentation within the boundaries of a given set of co-ordinates. By doing this, after the segmentation has been performed, any voxels selected outside the restricted region will be subtracted from the VOI. Therefore, the segmentation of a lesion can be confined within a specific area or segmentation of physiological uptake can be confined within a set of co-ordinates and then removed from a segmentation of disease. The latter technique is how the heart has been removed in Figure 2.25B, with the initial segmentation using a fixed 2.5 SUV threshold has included the heart and another segmentation constrained to a VOI placed around the heart, still using a fixed 2.5 SUV threshold, is used to segment the heart and remove it from the initial VOI.

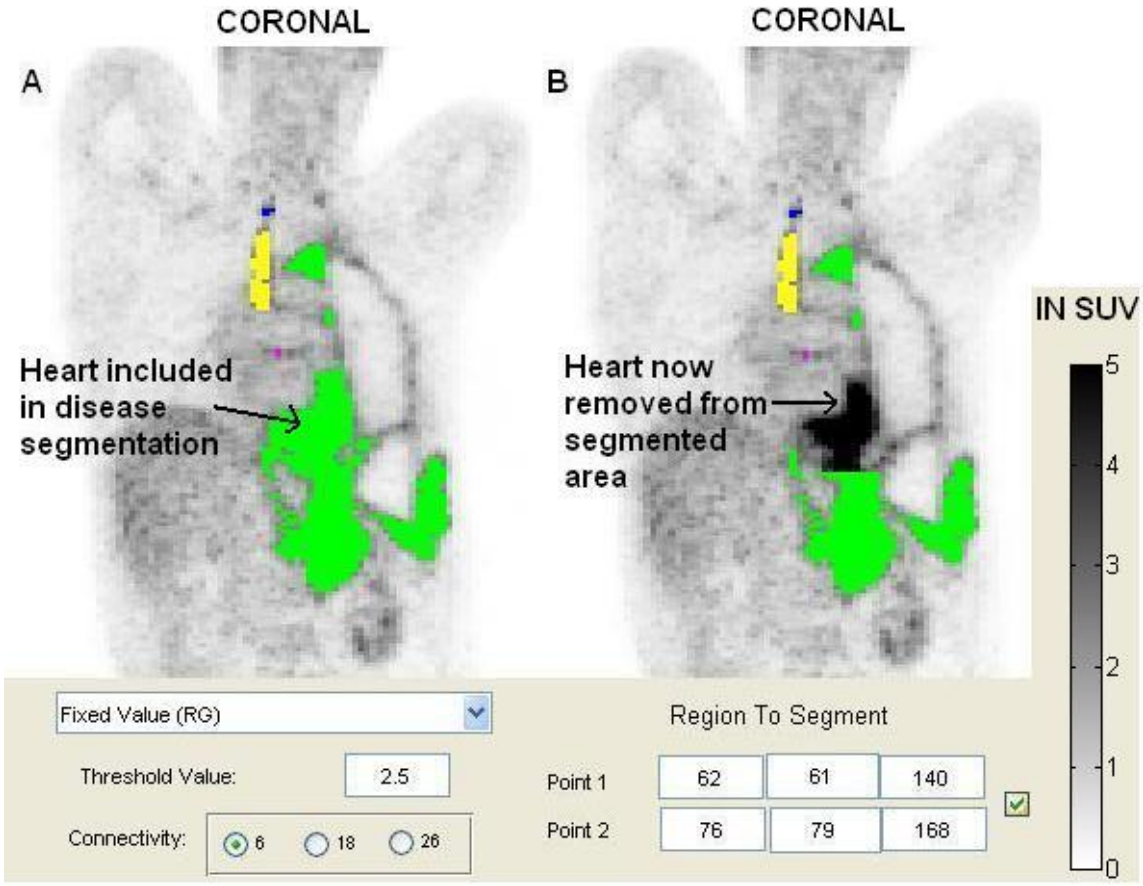


Figure 2.25: Visual Display of a Segmentation where the Heart is removed from Disease.

(A) User segments all the disease in the first image using fixed 2.5 SUV threshold but physiological uptake in the heart is included. (B) Physiological heart uptake is removed by using another fixed 2.5 SUV threshold restricted to the region between the co-ordinates 62-76 in the x-direction/coronal plane, 61-79 in the y-direction/sagittal plane and 140–168 in the z-direction/transverse plane. This segmentation of the heart is then removed from the initial segmentation so only disease is left.

2.3.10) Obtaining Background Regions in PETTRA

Obtaining a background region in PETTRA is important in many of the segmentation methods implemented in the software. There are two user input boxes for background mean and S.D. which are subsequently used in segmentation algorithms (Figure 2.24). These can be filled with values chosen by the user or, alternatively, the user can navigate to a voxel in the background

region and select ‘Calculate Background Region’ which autofills the input boxes with the mean and S.D. of the selected voxel and its 26-connected neighbours. If a larger background region is desired, the cubic volume segmentation can be used to obtain a volume in an area of background of a given size (this is described in more detail in 2.3.2). The mean and S.D. from this region can be used in the input boxes and, therefore, be used for segmentation.

2.3.11) Segmentation Methods not included in PETTRA

Segmentation methods discussed in 1.3.4 which have not been included in the PETTRA software include gradient based methods (Geets *et al*., 2007), dual active contours (El Naqa *et al*., 2007; Li *et al*., 2008), probability and possibility based methods (Hatt *et al*., 2009; Dewalle-Vignion *et al*., 2011), and AI methods (Sharif *et al*., 2010). In the majority of cases, the reasons why these segmentation methods have not been implemented is due to a combination of complexity and lack of testing on clinical data. While a lot of methods have potential to provide more accurate TVs than threshold based methods, they are at early stages of development and have not been fully tested on a range of clinical data. Most of the methods are more complex to program compared to threshold techniques and, therefore, these methods have been omitted from PETTRA. However, implementation of additional segmentation methods in the future is possible. Adaptive thresholding methods, discussed in 1.3.4 and 2.3.7, which rely on phantom measurements of T/B or *a priori* estimates of tumour size have also not been implemented in PETTRA (Jentzen *et al*., 2007; Van Dalen *et al*., 2007; Nehmeh *et al*., 2009). This is because there is no evidence to suggest they are more beneficial in comparison to the threshold techniques already implemented in the software and they rely on measurements which are unlikely to be available for many datasets when assessing response.

2.4) Quantification in PETTRA

2.4.1) Implementation of SUV

SUVs are generally regarded as the best, practical semi-quantitative measure to quantify PET images (Stahl *et al*., 2004; Nahmias and Wahl, 2008). Images are usually stored in intensities of Bq/cc, meaning there is no normalisation of tracer uptake with regards to radioactivity and patient volume e.g. body weight. Therefore, software must use patient weight, dose, radiopharmaceutical injection time and scanning start time to correct for radiotracer activity and patient volume to convert Bq/cc to SUV. These values are obtained from information from the image header file. Different image formats can use different header parameters and names, so coding a function which converts an image into SUV can be challenging. Even the same image format can store information differently depending on the institution and software used.

PETTRA's function for converting an image from Bq/cc to SUV starts by searching through the image file for radiopharmaceutical injection time, scan start time, administered dose to the patient, patient body weight, and isotope half life, slope and intercept on a logarithmic plot of the radioactive half life of the isotope. Without these seven parameters, SUV can not be calculated. If the function can not find these parameters in the header files it may ask the user to input the values manually. Knowing these inputs, the gap between the start of the scan and radiopharmaceutical injection time can be calculated and dose at the start of the scan can be derived. SUV can then be calculated using the following equation:

$$SUV = \frac{((Slope * Bq/ml) - Intercept) * (Patient\ Weight\ (kg) * 1000)}{Dose\ at\ Scanning\ Start\ Time}$$

[2.10]

2.4.2) Implementation of SUV Variations

As well as being able to convert images into SUV, PETTRA can also convert images into variations of SUV, including SUL, or SUV_{LBM}, and SUV_{BSA} (Zasadny and Wahl, 1993; Kim *et al*., 1994). Both are similar parameters to SUV and are calculated in the same way but instead of normalising by patient weight, SUL uses lean body mass (LBM) and SUV_{BSA} uses body surface area (BSA), otherwise formula [2.10] remains the same. Both are used throughout literature as better methods of normalisation than body weight (Graham *et al*., 2000; Hallett *et al*., 2004), and SUL is recommended as standard in PERCIST criteria (Wahl *et al*., 2009). Both SUL and SUV_{BSA} require patient height and weight, and SUL also uses the sex of the patient for calculation. These can be obtained from the image's header file like the other parameters. The equation for LBM is:

$$\text{If Male, } LBM = (1.10 * BW) - (120 * (BW / H)^2) * 1000$$
$$\text{If Female, } LBM = (1.07 * BW) - (148 * (BW / H)^2) * 1000 \quad [2.11]$$

where BW = body weight (kg) and H = Height (cm).

BSA is calculated as:

$$BSA = (0.007184 * BW^{0.425} * H^{0.725}) * 10000 \quad [2.12]$$

where BW = body weight (kg) and H = Height (cm). While LBM in SUL and patient weight in SUV are multiplied by 1000 to convert from kg to g, BSA is multiplied by 10000 to convert to cm^2/cc for calculation of SUV_{BSA}. In PETTRA, the default setting converts the image(s) into SUV, however, the user can toggle the setting to convert the values into SUL, SUV_{BSA} or their original Bq/cc values (Figure 2.26). SUV, SUL, and SUV_{BSA} values from PETTRA have all been compared with those from HERMES viewing system and they match perfectly.

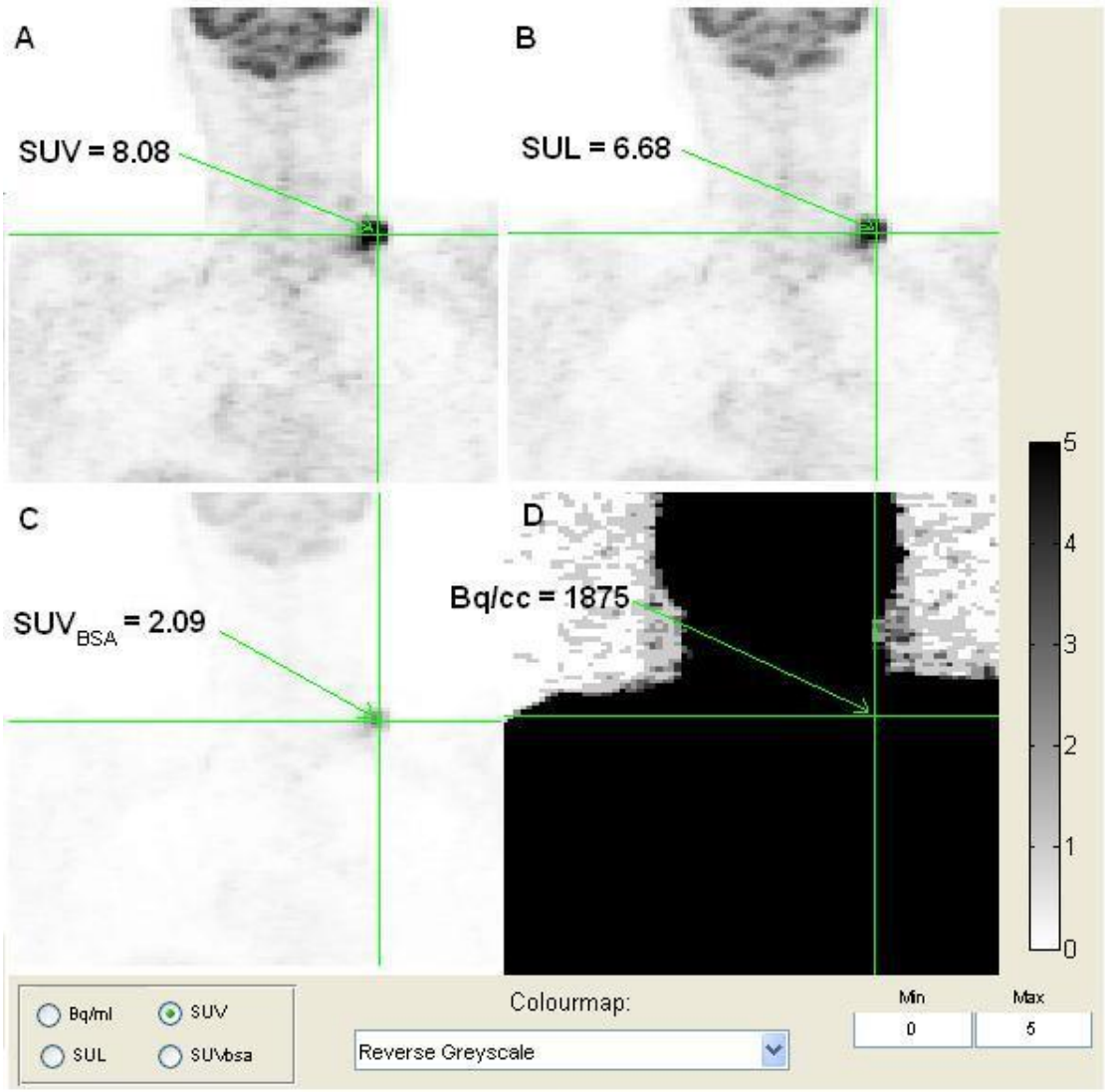


Figure 2.26: Different Units of Quantification in PETTRA

The same image displayed in (A) SUV, (B) SUL, (C) SUV_{BSA}, and (D) Bq/cc in PETTRA. The value of the highlighted voxel in the image is illustrated, as is the toggle for selecting what format the values are displayed in and the colour scale and range used for viewing the image.

2.4.3) Obtaining SUV_{max}

For assessing PET images, obtaining the SUV_{max} in an image or within a tumour is important in software such as PETTRA and there are buttons to find SUV_{max} in an image or within a VOI. The SUV_{max} is also given out in the display of details from a VOI.

2.4.4) Obtaining SUV_{peak}

SUV_{max} can be sensitive to noise in an image and so using SUV_{peak} is recommended in the literature (Wahl *et al*., 2009; Lodge *et al*., 2012). The methodology used to define SUV_{peak}, a small volume of the most intense uptake in a tumour, can cause results to vary considerably (Vanderhoek *et al*., 2012). PERCIST recommends that the SUV_{peak} is a $1cm^3$ sphere with a 1.2cm diameter centered on the hottest point in the tumour. The hottest point will typically contain the SUV_{max} but does not always. The implementation of SUV_{peak} into imaging software can be complicated depending on how it is achieved.

Images can be interpolated into a format where a sphere with a 1.2cm diameter and $1cm^3$ volume can be produced and the largest peak can be searched for throughout the image or VOI. This makes the task easier but also changes the original values of the PET image, something which is not advisable and which PETTRA refrains from. With no interpolation used, implementation of SUV_{peak} has to define a volume closest to $1cm^3$ with a 1.2cm diameter, if following PERCIST guidelines (Wahl *et al*., 2009). If a PET image has a voxel size of $97.80mm^3$ (5.47mm x 5.47mm x 3.27mm), then 10 voxels are needed to get as close to $1cm^3$ as possible, with a volume of $0.978cm^3$. If the PET image has voxel sizes of $71.85mm^3$ (4.69 x 4.69 x 3.27) and needs 14 voxels are needed to get as close as possible to a $1cm^3$, with a volume of $1.006cm^3$. This means that the shape of the SUV_{peak} in images with different voxel sizes is likely to be different. In addition, the shape constructed using different methodologies and software can result in different shapes, and therefore values, of SUV_{peak} as they will not be fully spherical.

For example, if a SUV_{peak} with a volume of $2mm^3$, rather than $1cm^3$, was required, only two voxels would be needed. These two voxels would be different values depending on which plane the neighbouring voxel was searched for in, i.e. a 2-voxel volume along the coronal plane will

yield a different result to one along the sagittal plane, to a different one along the transverse plane. Simple rectangular VOIs for 2-voxel (2 x 1 x 1), 4-voxel (2 x 2 x 1), 12-voxel (3 x 2 x 2), 16-voxel (4 x 2 x 2) and 18-voxel (3 x 3 x 2) SUV_{peak}-like measurements can be constructed with the difference in SUV_{peak} dependent upon which plane the VOI is in and results can vary by ~8% (0.3 SUV) (Table 2.1, based on the data of a group of mesothelioma patients in chapter 4.2). This may not seem significant but maximum differences of up to 38% (4.2 SUV) for 2-voxel VOIs and 24% (3.3 SUV) for 4-voxel and 16-voxel VOIs are possible which could have a considerable effect on the reproducibility and variability of SUV_{peak} between studies. The lowest differences are seen in volumes with a higher number of voxels and with more symmetrical shapes (the 16-voxel SUV_{peak} has larger differences, ~10%, but this is because the cubic volume is more elongated with dimensions of 4 x 2 x 2).

The data herein show that even at a fundamental level, SUV_{peak} is prone to variation and matters are further complicated when different shapes or connected voxels can be used, as well as different directions. For example, if a spherical SUV_{peak} is constructed of $1cm^3$ (10 voxels) with a 1.2cm diameter (2 voxels in coronal or sagittal plane or between 3-4 voxels in the transverse plane) for this data, several options are possible as potential shapes that can be used (Figure 2.27).

SUV_{peak} Type	SUV_{peak} Value			
	Mean Diff	**Mean Diff (%)**	**Max Diff**	**Max Diff (%)**
2-voxel SUV_{peak} (2 by 1 by 1)	0.682 (0.831)	7.55 (7.56)	4.198	37.39
4-voxel SUV_{peak} (2 by 2 by 1)	0.550 (0.578)	6.29 (6.18)	2.989	24.77
12-voxel SUV_{peak} (3 by 2 by 2)	0.396 (0.476)	4.62 (4.19)	2.514	19.82
16-voxel SUV_{peak} (4 by 2 by 2)	0.612 (0.701)	7.33 (5.74)	3.288	24.10
18-voxel SUV_{peak} (3 by 3 by 2)	0.360 (0.461)	3.90 (3.98)	2.106	16.41
ALL	0.520 (0.609)	5.94 (5.53)	4.198	37.39

Table 2.1: Mean and Maximum Differences of Rectangular SUV_{peak} in Different Planes

Mean and maximum difference of SUV_{peak} measurement when taken in different planes in all three directions (i.e. x-plane value – y-plane value, x-plane value – z-plane value and y-plane value – z-plane value) over 27 images with a voxel size of 97.80mm^3in a dataset of mesothelioma patients (see chapter 4.2 for details). Absolute and percentage differences are shown. S.D. (Standard Deviation) in brackets. Diff = Difference.

SUV_{peak} in PETTRA was implemented by joining the 6-connected neighbours (then 18-connected and then 26-connected) in a systematic and logical order, with voxels added on the coronal plane, then the sagittal plane and finally the transverse plane. Using 10 voxels for SUV_{peak}, PETTRA would use the shape in Figure 2.27A. However, more symmetrical and sphere-like shapes in Figures 2.27B and 2.27C could represent the SUV_{peak} of a lesion more accurately. However, these methods would have a length/'diameter' of ~16mm (3 x 5.47mm), while the shape in Figure 2.27D would be closer to the PERCIST recommended 1.2cm diameter, at 10.94mm (2 x 5.47mm). Similarly to the difference in direction, the difference in shape used can affect the SUV_{peak} in the region of 3-4% (Table 2.2). All the different shapes and planes could be used to search for the highest region. However, this is more computationally exhaustive and is likely to be different depending on the exact methodology and software.

To create more spherical VOIs with diameters closer to 1.2cm, taking partial voxels can be used rather than using whole voxels, e.g. using half voxels to produce more spherical like VOIs while using half the intensity of the whole voxel to calculate the intensity. However, computing a

SUV_{peak} using such a methodology is more intricate. While this was not investigated in this study, it would be interesting to discover the affect this technique would have on SUV_{peak}. It is worth noting that this is another factor which could cause variation in SUV_{peak}.

	Difference in SUV_{peak} Value	
	Diff (SUV)	**Diff (%)**
Mean	0.329	3.19
S.D.	0.483	3.55
Min	0.002	0.06
Max	2.626	17.74

Table 2.2: Mean Difference of SUV_{peak} using Different Shapes of 10 Voxels

Mean difference in SUV_{peak} using shapes B, C and D in Figure 2.27 over all 27 images with a voxel size of $97.80mm^3$. Absolute and percentage differences are shown. S.D. = Standard Deviation. Diff = Difference.

There are also significant differences in SUV_{peak} values depending on the size of the VOI and whether it is in 2-D or 3-D (Vanderhoek *et al*., 2012), however, even if a fixed 3-D volume is taken this does not always mean the results will be the same as has been shown with the results in Tables 2.1 and 2.2. While these variations need to be considered, as long as the method uses the same volume and methodology it is unlikely to vary enough to cause significant differences. For example, over the 27 images with voxel sizes of $97.80mm^3$, the default PETTRA SUV_{peak} shape A correlated extremely well with shapes B, C and D (Figure 2.27), with Pearson Correlation Coefficients (pcc) of 0.998, 0.996 and 0.990 respectively (p-value > 0.0001).

In PETTRA, SUV_{peak} in a segmented VOI is obtained by calculating the SUV_{peak} for each voxel in the VOI with the highest value representing the SUV_{peak} in the VOI. This does mean that the SUV_{peak} may include voxels outside the segmented VOI, however, this would be unlikely, especially in large TVs. SUV_{peak} was also obtained from HERMES software, which tries to

follow the PERCIST recommendations but often takes a lower volume, presumably to take into account the SUV_{peak} diameter measure.

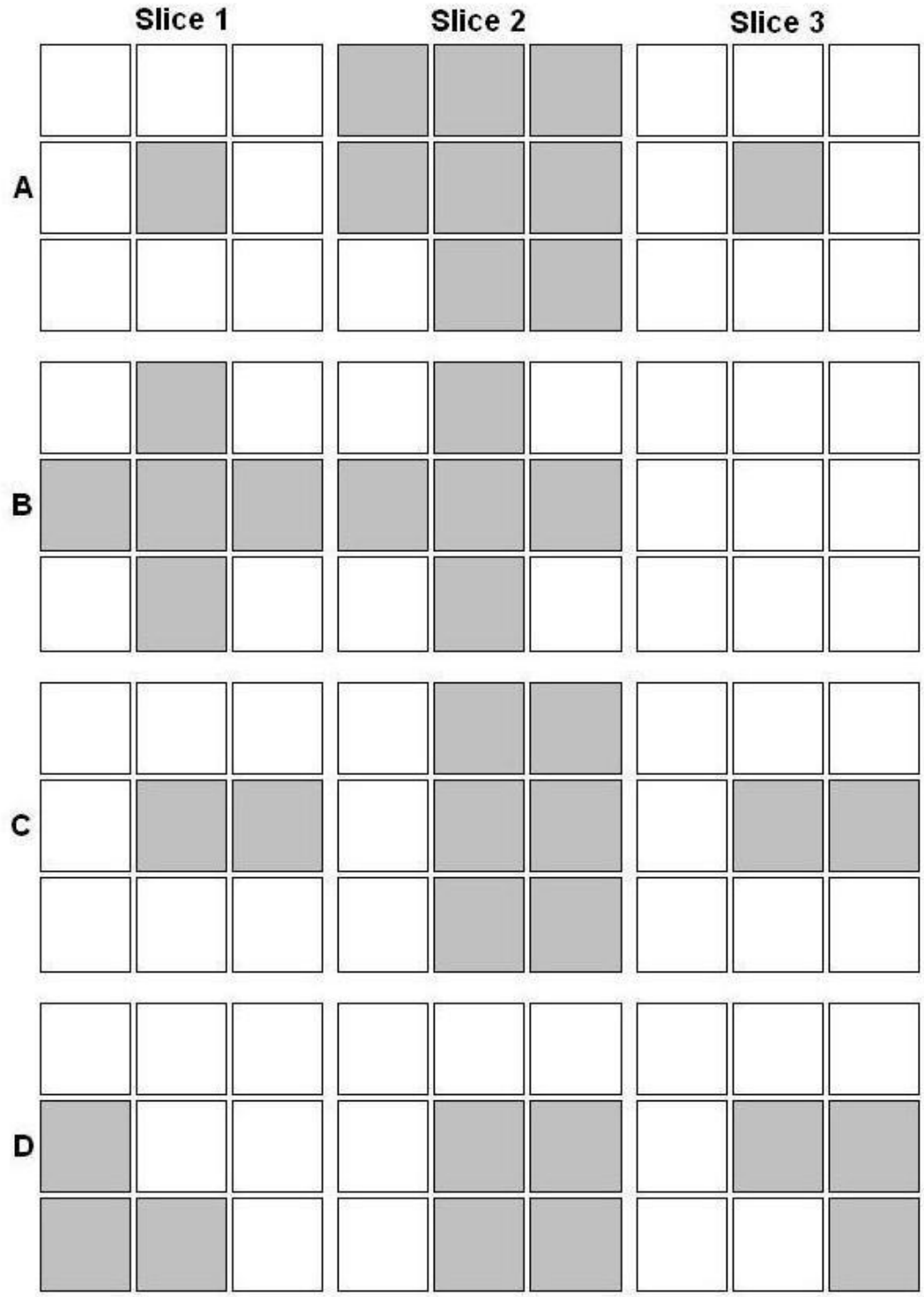
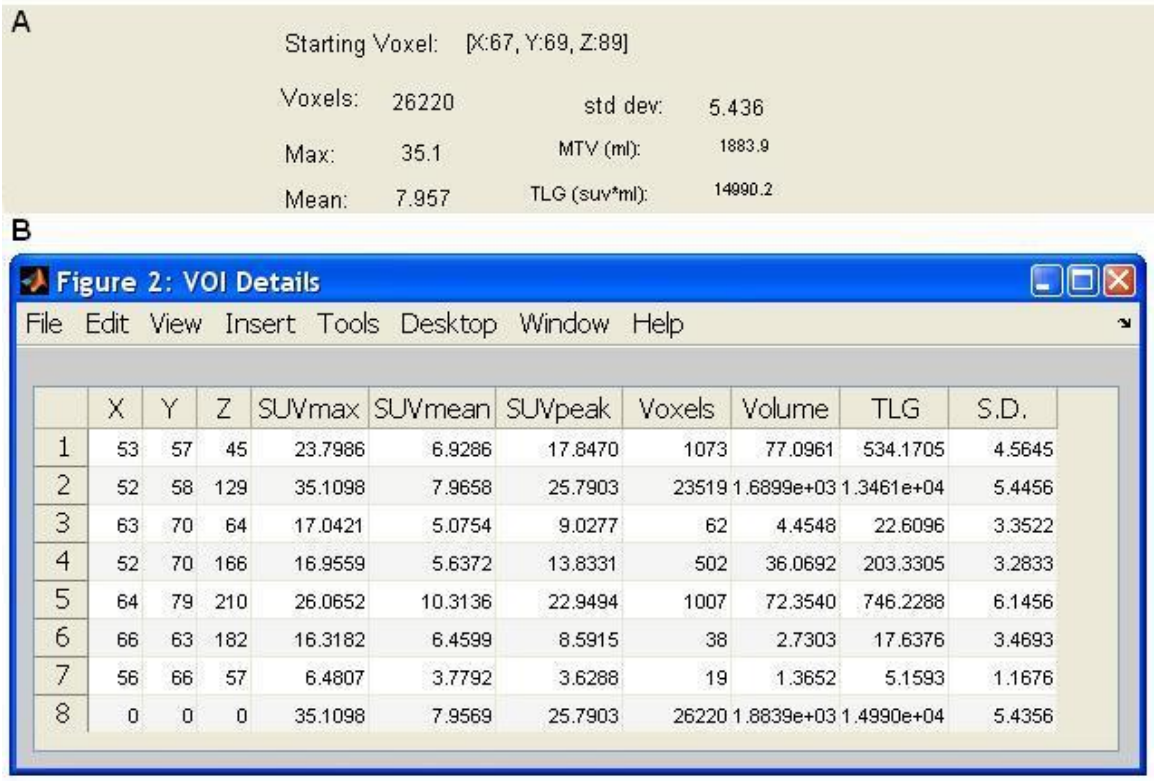


Figure 2.27: Potential Shapes for SUV_{peak} using 10 Voxels

2.4.5) Obtaining SUV_{mean}, Tumour Volume and Total Lesion Glycolysis

Once disease has been segmented, parameters such as SUV_{mean}, TV, TLG and S.D. can all be obtained from a tumour and are useful in assessing response along with SUV_{max} and SUV_{peak} parameters. PETTRA will automatically display the SUV_{max}, SUV_{mean} and volume parameters for the combined segmented areas in the viewer and in addition will show all these parameters, and SUV_{peak}, for each VOI in a table upon the request of the user (Figure 2.28).



	X	Y	Z	SUVmax	SUVmean	SUVpeak	Voxels	Volume	TLG	S.D.
1	53	57	45	23.7986	6.9286	17.8470	1073	77.0961	534.1705	4.5645
2	52	58	129	35.1098	7.9658	25.7903	23519	1.6899e+03	1.3461e+04	5.4456
3	63	70	64	17.0421	5.0754	9.0277	62	4.4548	22.6096	3.3522
4	52	70	166	16.9559	5.6372	13.8331	502	36.0692	203.3305	3.2833
5	64	79	210	26.0652	10.3136	22.9494	1007	72.3540	746.2288	6.1456
6	66	63	182	16.3182	6.4599	8.5915	38	2.7303	17.6376	3.4693
7	56	66	57	6.4807	3.7792	3.6288	19	1.3652	5.1593	1.1676
8	0	0	0	35.1098	7.9569	25.7903	26220	1.8839e+03	1.4990e+04	5.4356

Figure 2.28: Visual Display of VOI Statistics in PETTRA

Details of VOIs segmented by the user are automatically displayed on the PETTRA GUI, including the starting voxel of the last segmentation (A). A pop-up table is displayed after 'VOI Details' is selected which shows full details of each individual segmentation in the image and the combined total (B). Each row represents a different individual segmentation with the last row being the SUV_{max} and SUV_{peak} over all the segmentations and the combined total of SUV_{mean}, number of voxels, volume, TLG and S.D. The X, Y and Z co-ordinates on each row represent a voxel from within the segmented region.

The volume of each VOI is listed in ml and is the result of the number of voxels multiplied by the voxel size, obtained from information in the header file of the image. TLG is the product of the SUV_{mean} and the volume (SUV * ml, or g).

2.4.6) Implementation of Intensity Volume Histograms and Related Statistics

As discussed in 1.3.5, histograms can provide more information about a VOI than simply SUV_{max} and SUV_{peak}. PETTRA has implemented a 'Statistics' button which brings up a new GUI which presents the user with a wealth of graphs and parameters that can be used for identifying response to therapy including several histograms which can be a useful visual aid and which parameters can be extracted from. Firstly, typical histograms plotting the number of voxels against intensity, with different binning of voxel intensities, are displayed (Figure 2.29).

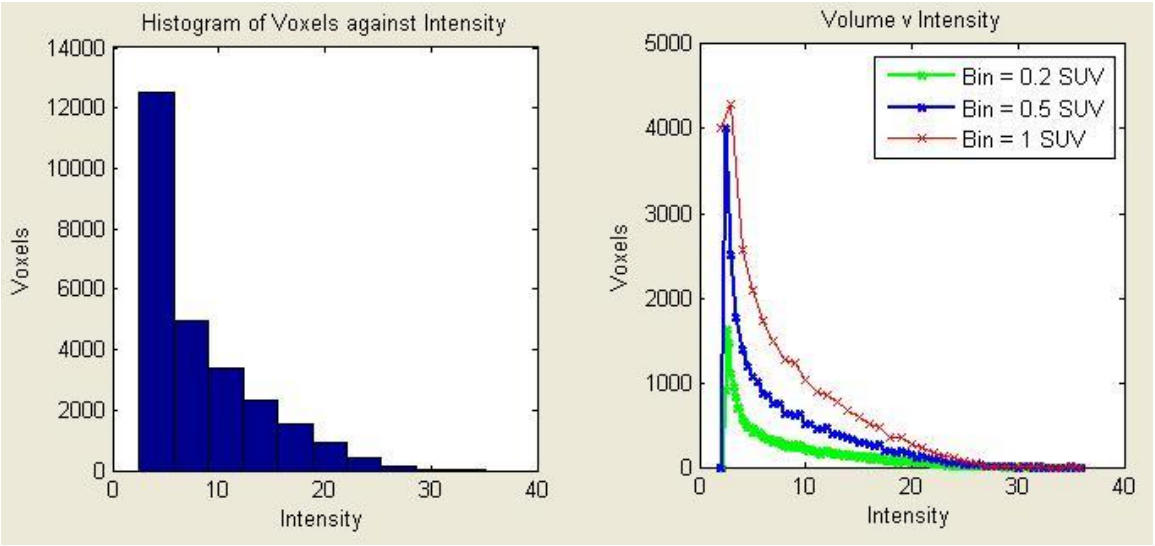


Figure 2.29: Histograms of Number of Voxels against Intensity in a Segmented VOI

(A) MATLAB constructed histogram of voxels against intensity and (B) a customised version of the histogram with data binned into units of 0.2, 0.5 and 1 SUV.

The intensity volume histogram (IVH) is a modified version of the histogram which plots volume as a function of image intensity over a given intensity. A number of metrics can be extracted from the IVH, the six used in studies looking at response are I_{10}, I_{90}, $I_{10\text{-}90}$, V_{10}, V_{90} and $V_{10\text{-}90}$ (El Naqa *et al*., 2009a; El Naqa *et al*., 2009b). I_x is defined as the 'minimum intensity to x% highest intensity volume' while V_x is defined as the '% volume having at least x% intensity value' (El Naqa *et al*., 2009a). These have been implemented in PETTRA as I_{10} (or I_{90}) is the minimum intensity in the highest 10% (or 90%) of the volume subtracted from the minimum intensity in the VOI, while V_{10} (or V_{90}) is the volume in the VOI having at least 10% (or 90%) of the SUV_{max}. I_{10} and V_{10} are therefore calculated as:

$$I_{10} = SUV_{min} \text{ in the highest 10\% of VOI} - SUV_{min} \quad [2.13]$$

$$V_{10} = \frac{(\text{Voxels in VOI} > (0.1 * SUV_{max}))}{\text{No of Voxels in VOI}} * 100 \quad [2.14]$$

To show the principle of these equations, an example from the dataset investigated in chapter 5 is shown. Table 2.3 lists the intensity of the 17 voxels in a TV on a response scan. For this example, I_{10} would be the minimum intensity of 4.6 and 4.5, as just two voxels would make up 10% of the volume. These subtracted from the minimum intensity of 2.6 would give a I_{10} of 1.9. V_{10} would result in a volume of 100% because all the voxels would be over 10% of the maximum intensity of 4.5. However, V_{90} would give a volume of 17.64% as there are three voxels greater than 90% of the SUV_{max} (0.9 * 4.6 = 4.14) and three divided by the number of voxels in the VOI (17) gives 17.64%.

4.6	4.3	3.9	3.7	3.3	3.1	3.0	2.7	2.6
4.5	3.9	3.8	3.5	3.1	3.0	2.9	2.6	

Table 2.3: List of Voxel Intensities from a Tumour Volume in a Response Scan

The concept of these IVH parameters is that they describe more about the distribution and intensity of data than SUV_{max} and SUV_{peak} and take into account the tumour intensity unlike TV. For example, I_{10} is taking into account a specific intensity range which does not involve the SUV_{max} or the top 10% of intensities in the image, but the intensity range after they are removed. This eliminates the possibility of noise affecting these values considerably and estimates whether the tumour is still very active if the highest values in the tumour are removed. If a tumour still has a high intensity after the highest intensities have been removed, it is likely to have more, intense cancer cells and, theoretically, be more immune to therapy and less likely to respond. The IVH and related statistics, including volumetric data for a VOI, are displayed in PETTRA as shown in Figure 2.30.

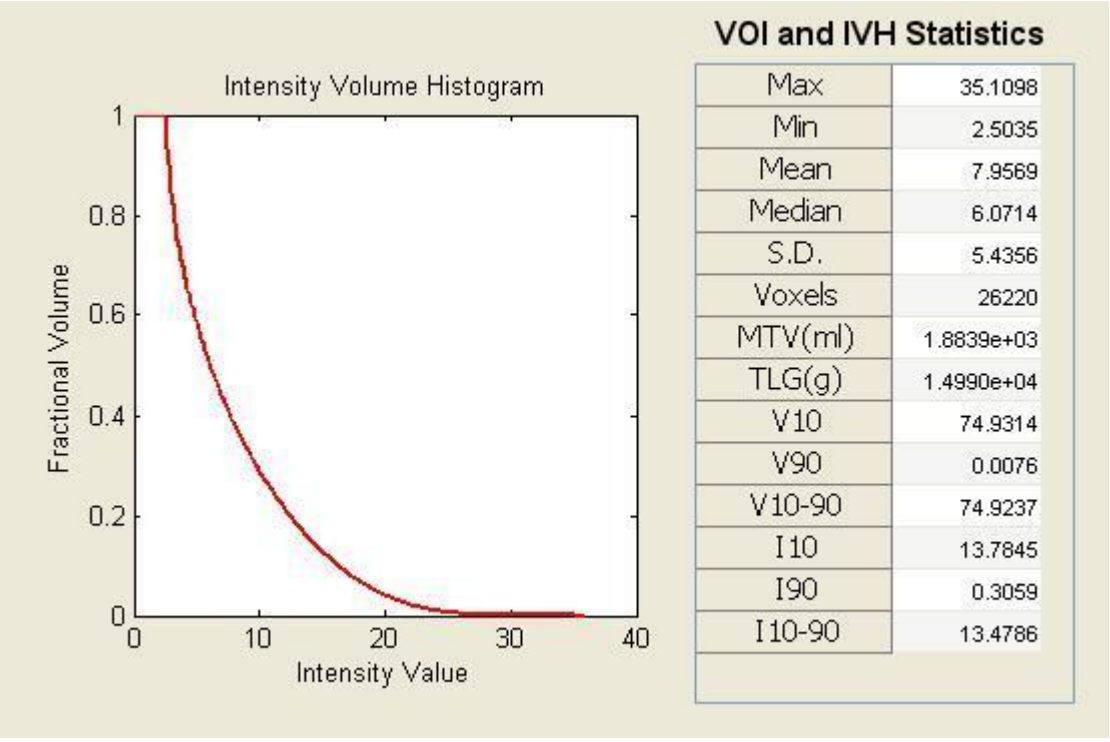


Figure 2.30: Intensity Volume Histogram and Related Statistics for a Segmented VOI

The IVH is displayed on the left while statistics on the right show related parameters to the IVH and statistics taken from the VOI.

3) Registration of Pre- and Post- Therapy PET/CT Scans for Assessing Response to Therapy

3.1) Introduction to Registration of Pre- and Post- Therapy PET/CT Scans

3.1.1) Motivation for Registration of Pre- and Post- Therapy PET/CT Scans

There are several reasons for obtaining registered pre- and post- therapy PET/CT images, particularly for assessing response to therapy. Qualitatively, image registration can allow an observer to see lesions in both scans in the same anatomical space, potentially aiding the interpretation of pre- and post- therapy images and the determined response. Quantitatively, accurate registration can allow computation of subtraction images of pre- and post- therapy images from which it may be possible to obtain new parameters. Tumour volumes (TV) from pre- therapy scans can be used on post- therapy scans to analyse changes in the size, shape and texture of lesions, potentially aiding the development of volumetric measures such as total lesion glycolysis (TLG). If registration is accurate enough, new methods of quantitative response assessment employing voxel analysis techniques, such as statistical parametric mapping, could be used, as they have been in studies of the brain (Jeong *et al*., 2005; Lee *et al*., 2008; Amorim *et al*., 2010; Hsieh *et al*., 2012). These methods have the potential to improve response assessment and predict the success of therapy. Therefore, a robust and accurate method for registering pre- and post- therapy PET/CT studies could be of significant benefit. In this chapter, three methods for registering pre- and post- therapy PET/CT images are investigated and evaluated in respect of their accuracy and robustness. The effect of the registration methods on SUVs and TVs will also be studied.

3.1.2) Registration of Pre- and Post- Therapy PET/CT Scans

PET images are far from ideal to register, as they have low spatial resolution meaning that the best possible registration can still only be within the region of 4-6mm and the metabolic nature of imaging means that anatomical (or other) landmarks cannot be selected for use in registration easily. Intensities in the image are likely to change significantly to mark a response to treatment ruling out the possibility of using them for registration. For these reasons, registering one PET scan to another becomes problematic. Registration of pre- and post- therapy PET images has been attempted using a non-rigid algorithm which uses position dependent rigidity to stop areas of high tracer uptake changing shape or intensity (De Moor *et al*., 2006). Registration was found to be accurate and made visual therapy response assessment easier and faster, however, there was no thorough evaluation into how accurate the registration was or how it may help assess response assessment other than promising visual images and clinician feedback.

While PET-PET registration is a plausible method for registering PET/CT scans, using the CT component for registration appears to be a more practical, easier and justifiable method of registration. CT images have a high resolution and are not as likely to change dramatically over time, particularly when compared to PET where drastic changes in tumour uptake are expected shortly after therapy. Although changes in tumour size and density are seen on CT images, they are not usually as sizeable after a few weeks of therapy (Wieder *et al*., 2005), and therefore, registering pre- and post- therapy CT images and applying the resulting transformations to the respective PET images appears to be a more suitable method. A similar methodology was used to register PET/CT studies of lung tumours, using a multilevel nonlinear registration to align pre- and post- therapy CT data to be used in the registration of PET images (Ouksili *et al*., 2007). This study showed promising results, although it was only tested on synthetic data, and not clinical data.

Most other studies investigating PET/CT registration opt for rigid registration methods (Necib *et al*., 2011; Zsoter *et al*., 2011). Rigid registration does not use non-rigid deformations to register tumours, therefore eliminating the possibility that the size and shape of a tumour on the registered post-therapy image will be changed drastically in an attempt to match it to the pre-therapy image. Changes in tumour size from pre- to post- therapy CT images, the potential effect of these changes on registration and the affect of all this on the transformed PET images are yet to be full studied. Despite these potential issues, non-rigid registration is likely to register images more accurately than rigid registration and has been shown to be very accurate in aligning head and neck PET/CT scans for radiotherapy planning (Ireland *et al*., 2007). However, the most recent studies on the subject use rigid registration techniques. A hextuple registration method using a unique similarity measure showed good accuracy and speed in registering simulated pre- and post- therapy PET/CT scans but this has not yet been applied to clinical data (Zsoter *et al*., 2011). Necib *et al*. (2011) computed parametric images by registering pre- and post- therapy PET/CT scans and obtained promising results in terms of predicting response, suggesting registration must have been successful for this methodology to work, however, the accuracy of registration was not thoroughly investigated (Necib *et al*., 2011).

3.1.3) Registration of PET and CT Scans

PET/CT scanners allow acquisition of both modalities within the same image space allowing automatic co-registration. However, there are limitations due to differences in imaging time, which can result in errors in registration due to patient movement and breathing. During a PET/CT scan, the CT scan is acquired at the beginning of the imaging protocol and the more time consuming PET after. This means the CT image is from a relatively fixed position within the breathing cycle and of a still patient, while the PET image will be over the whole breathing cycle and is likely to be affected by a patient's movement on the bed. These differences between the acquisition of the PET and CT scans can cause errors in registration (Veit *et al*., 2006). If

common acquisition and reconstruction methods are used, images should still be very well registered in space and in time. PET and CT images from PET/CT scanners can be registered with an unprecedented accuracy compared with separate imaging (Coleman *et al*., 2005; Weigert *et al*., 2008).

In some PET/CT scans there is the possibility of non-rigid misalignment artefacts due to respiration (Shekhar *et al*., 2005), although these can be minimised to dimensions comparable to the spatial resolution of most PET scanners (Goerres *et al*., 2002a). Statistically significant differences between PET and CT images are common but are of a modest nature and are usually in the limits of PET resolution (Nakamoto *et al*., 2003). A study measuring distance between fiducial markers on concurrent PET and CT scans found differences between 1.6mm to 3.1mm (Somer *et al*., 2007), within the range of PET resolution. Therefore, it appears acceptable to apply transformations obtained from registering pre- and post- therapy CT scans to concurrent pre- and post- therapy PET scans. However, it should be noted that differences in PET and CT acquisition can cause misalignment and registration of concurrent PET and CT scans could be used to improve registration of pre- and post- therapy PET/CT studies further.

3.2) Methodology for Registering Pre- and Post- Therapy PET Scans

3.2.1) Proposed Registration Algorithms for Registering PET Images

CT-CT registrations are intra-modality and intra-subject class problems, meaning that the modality and subject remain the same with the only differing factor being the time of the scan. This makes voxel intensity based similarity measures for registration a favourable and fast measure to use. The sum of squared differences (SSD) is a measure which uses the difference, or lack of, from one image to another in terms of intensity values to guide registration. This is the

most common measure for registering images of the same modality and has been used for registering serial MR scans (Hill *et al*., 2001). Other intensity based similarity measures such as cross correlation (CC) and normalised mutual information (NMI) will be tested but are more suited to inter-modality registration problems i.e. scans from different modalities. Other registration techniques, such as using landmark or surface/edge-based measures, have been disregarded because there are no fiducial markers in the images. Identifying anatomical landmarks would be time consuming and not as accurate as the aforementioned methods as intensity values across the whole CT image hold the most information and rarely change drastically.

There has been limited research on longitudinal pre- and post- therapy PET-PET or CT-CT registration. Designing new methods and algorithms would be wasteful when current techniques and software may be adequate. For this reason, the Image Registration Toolkit (IRTK), which provides an environment in which many types of medical images can be registered to one another using both rigid and non-rigid algorithms, was used for registration (Rueckert *et al*., 1999; Studholme *et al*., 1999; Schnabel *et al*., 2001). IRTK has been shown to be one of the best nonlinear deformation registration methods for registering MRI brain images (Klein *et al.*, 2009). The IRTK library is flexible and allows parameters to be modified to ensure the best registration for a particular task, in this instance, registering longitudinal pre- and post- therapy CT scans of lymphoma patients using rigid and non-rigid techniques.

Both customised rigid and non-rigid IRTK registration algorithms were used to register pre- and post- therapy CT studies with parameters such as resolution levels, blurring, number of iterations, step size, step length, interpolation mode and optimisation method experimented with to find robust and accurate registration methods suitable for the purpose of registering pre- and post- therapy CT scans. Initial experimentation took place on five patients, from a dataset of 20

patients with lymphoma. The full 20 patient dataset was used to validate the registration methods in a more thorough assessment. In the initial experimentation, visual interpretation of pre-therapy and registered post- therapy images, both overlaid in greyscale and in the form of subtraction images, was used to assess which parameters provide the most accurate, efficient and robust algorithms for rigid and non-rigid registration. Subtraction images subtract the intensity values of one image from the other allowing the observer to see differences between them and the changes in intensity. More accurately registered post-therapy images are likely to produce subtraction images with less difference in intensity values.

Patient alignment between pre- and post- therapy scans is usually relatively close in the coronal and sagittal planes as the patient is placed in the same position along the width of the bed. However, the misalignment in the transverse plane is often much greater as the patient positioning along the length of the bed is more variable. Patients should be aligned within 3cm, in either direction, of a marker used to guide patient positioning along the length of the bed to ensure that misalignment between scans is no larger than this. This is the case in the dataset used herein acquired at the St Thomas' PET Centre but also, a similar standard of positioning should be used in other institutes. This still means PET-CT images can be misaligned by up to 6cm in the transverse plane before any registration is performed. Initial experimentation of changing parameters in registration algorithms found large misalignments of this degree to be problematic causing significant modifications of IRTK parameters. A method involving calculating the centre of mass (COM) of each scan and aligning these in image space showed promise in aligning images in a computationally fast manner and was tested along with IRTK rigid and non-rigid methods.

3.2.2) Centre of Mass Registration Algorithm

In an attempt to register images initially in a computationally efficient way, a COM registration method was implemented to try and perform a simple but very fast alignment. The COM of an image, I, is defined as:

$$\text{COM of I} = \frac{\sum m_i p_i}{\sum m_i} \qquad [3.1]$$

where p_i = position of each voxel and m_i = mass of each voxel.

Calculating COM of both pre- and post-therapy images in all three axes, the two can be aligned approximately by converging them together in all three planes. This was achieved using MatLab®. The difference in COM in each plane was calculated in voxels and then converted to mm so it could be corrected for on IRTK R-view software, a visual viewing tool for looking at registered images. This technique is very efficient, taking ~5-20s to calculate the COM on the two images and can potentially give a good initial starting point for finer algorithms to start from.

3.2.3) Rigid Registration Algorithm

Default IRTK input parameters for rigid registration assume the two images are already aligned relatively well, using low step lengths i.e. small movements in the region of mm rather than cms, to try and match the source image to the target image. Due to the possibility of considerable original misalignment, particularly in the transverse plane, input parameters were modified to allow successful alignment of all images. The customised algorithm uses three resolution levels with the first using very large step lengths and a high number of iterations to ensure the images are aligned to within a few cm. This allows even the most misaligned images to converge to an extent where finer registration can take place, increasing the robustness of the algorithm.

SSD is used as a voxel similarity measure, gradient descent as an optimisation method and default values for blurring and resolution of the images were kept at 1.6mm and 3.2mm respectively at the finest registration level. The blurring and resolution parameters are set to dictate to what degree the images are blurred defining the amount of detail in the images that stands out, and to set the voxel size of the images during the registration process. At the most coarse registration level, blurring and resolution are set to 6.5mm and 13mm respectively but with a large step length, number of steps and iterations to ensure convergence from poorly aligned pre- and post-therapy scans. The algorithm uses linear interpolation, as it is balances accuracy and speed (using more complex methods of interpolation at least tripled the time taken for registration). Nearest neighbour interpolation is not ideal in terms of getting a more detailed transformation and more complicated sinc and cubic methods increase computation time greatly with little significant difference in registration accuracy when viewing subtraction images.

3.2.4) Non-Rigid Registration Algorithm

Rather than use a non-rigid method that starts on unregistered images, it is more appropriate to use a rigid registration transformation as a starting point for a non-rigid method. The IRTK non-rigid method uses the IRTK rigid registration transformation as a starting point due to the considerable initial misalignment that can be found between some pre- and post- therapy scans. While the deformations applied by the non-rigid algorithm are likely to have an effect on the accuracy of registration at a fine level, they are unlikely to change transformations at the beginning of registration when initially aligning two images cm apart, and so applying them after the rigid algorithm appears to be a logical method. The IRTK non-rigid algorithm uses three resolution levels like the rigid registration algorithm, but with shorter step lengths, number of steps and number of iterations in an attempt to achieve finer alignment, all set to default IRTK settings. Like the IRTK rigid algorithm it uses SSD as a similarity measure and gradient descent as an optimisation method. A typical IRTK non-rigid registration takes ~1h to run but this is

variable depending on the exact size of the images/subvolumes being registered. While rigid registration allows only translations and rotations to align images, non-rigid algorithms perform affine registration changes, such as scaling and shearing of images so lengths and angles are not preserved, and non-linear deformations to register the source image to fit the target. The IRTK non-rigid method uses free-form deformation based on b-splines to transform images (Rueckert *et al.*, 1999).

Non-rigid registration is likely to improve registration accuracy of images, although which method will be better for looking at response to therapy is unclear as non-rigid registrations could cause unwanted changes to tumours on PET images. This will be an important factor in this study and is evaluated in 3.4. The non-rigid method takes between 45-111min (mean: 80min), depending on the size of the image or subvolume, compared to ~3-4min for the rigid method in comparison (on a Lenovo Thinkpad R61 with 3GB of RAM and an Intel® Core™ 2 Duo CPU T8100 @ 2.10GHz).

3.2.5) Dataset for Evaluating Registrations

All three registration algorithms were performed on 20 pre- and post- therapy PET/CT studies on patients with lymphoma (11 male, 9 female; median age, 30 years; age range, 18-73 years). The time between pre- and post- therapy scans in this study was approximately 3 and a half months (median, 96; days, range, 55-263 days). CT image dimensions were 512 x 512 x 223, 267 or 311 (with voxel sizes of 0.98mm x 0.98mm x 3.27mm) while PET image dimensions were 128 x 128 x 223, 267 or 311 (with voxel sizes of 4.69mm x 4.69mm x 3.27mm). In the transverse plane, the number of slices was 267 slices 25 times, 223 slices 13 times and 311 slices twice over the 40 images from the 20 datasets. All PET images underwent AC. Injected activity of ^{18}F-FDG for each PET scan ranged from 294-377 MBq. Both PET and CT scans used free breathing protocols and CT scans were low dose, non-contrast enhanced scans. Images were acquired on

GE Discovery ST or GE Discovery VCT PET scanners (Waukesha, WI). Response to therapy on CT only was measured by a radiologist using the IWC (Cheson *et al*., 1999). PET/CT response was scored by two physicians from the PET Imaging Centre at St Thomas using the Deauville criteria (Meignan *et al*., 2009).

The three registration algorithms were tested on subvolumes of each image, selected from below the pelvis to the base of the skull in the transverse plane, the complete torso in the coronal plane (with arms removed) and from the front to the back of the thorax in the sagittal plane. This subvolume was chosen as it was thought that it would allow registration to take place on a specific area of the body, as would be expected in research scenarios when registering just diseased areas of the patient. Equally, this area should give enough information for both visual and landmark analysis. On average, volumes were 335 x 230 x 630mm. Subvolumes were used for performing all registrations and for visual and landmark analysis. It is hoped that the algorithms used will work on both smaller and larger VOIs and even whole body images as well. However, for this study it was important to keep analysis to an area large enough to assess registration success while being small enough to ensure it was a practical investigation and computationally efficient. All the registration methods use CT data to perform registrations and the resulting transformations are then applied to corresponding PET images with the assumption that there was no, or limited, patient movement between the PET and CT acquisition. No noticeable misregistrations between PET and CT were found during inspection of the registrations.

3.2.6) Reason for Chosen Registration Evaluation Methods

Analysing image registration algorithms can be difficult to achieve, not least because it can be challenging to differentiate between registration inaccuracies and actual physical differences between images (Zitova *et al*., 2003). Qualitative assessment, where experienced image analysts

assess the images visually, can be a valid method but there are issues with how much misalignment can be detected by human evaluation. Wong *et al*. (1997) investigating human perception of registration accuracy in 3-D brain ^{18}F-FDG PET to MRI images showed five observers could all detect translational misregistration of up to 2mm along the x- and y- axes, and 3mm in the z-axis (Wong *et al*., 1997). Rotational misregistration could be detectable by all observers by up to 2° in z-axis, 3-4° in the y-axis, and 2° in the x-axis for positive rotation and 4° for negative rotation. A similar experiment on MR and CT brain images showed that differences of >2mm are usually spotted by experienced observers (Fitzpatrick *et al*., 1998a). This evidence suggests that visual assessment can be a worthwhile method of testing registration algorithms. There are clear differences between the degrees of detection that can be made between whole body CT-CT images in comparison to MRI brain images but it is still likely that any translations of several mm or significant rotation will be noticed.

Quantitative methods of assessing registration accuracy are less subjective and provide a result which can quantitatively estimate how accurate registration is. One method of quantitatively assessing registration accuracy is to use a voxel-based similarity measure to calculate how well aligned images are with each other. As this technique has been used to register images, it would be expected to show excellent results, but it is a circular argument to prove the method works using the same technique used to achieve the registration in the first place. An alternative method is to assess how close fiducial or anatomical landmarks are to each other on the two images by calculating the distance between them. Fiducial markers can be placed in or around a patient during scans. While this is ideal for quantitative assessment, fiducial markers are not used routinely in a clinical environment so the dataset does not have such markers available. Rather than using artificially placed fiducial markers, anatomy in images can be used. An anatomical landmark registration assessment is more practical but not ideal as there are likely to be differences in where markers are placed on images, for example, due to intra- and inter-

observer variability. Factors including patient breathing, movement and positioning can potentially have an impact on the accuracy of an observer locating these markers.

Although using anatomical landmarks has some operator variability, it can be an excellent measure of registration as it can test whether the algorithm meets the clinical needs of the observer (Hutton *et al*., 2002). Reliably identifying anatomical landmarks requires a high resolution of anatomy so while landmark analysis on CT images is plausible, on low resolution functional PET images, this sort of analysis is not really possible. Non-rigid algorithms are harder to assess than rigid ones as the extent of non-rigid misalignment can be greatly different from case to case, therefore no single validation method can be used confidently (Hutton and Braun, 2003). Despite this visual assessment has been proven to be successful at validating and comparing non-rigid registration and there is no evidence to suggest that quantitative methods such as anatomical landmark measures would not provide useful analyse (Meyer *et al*., 1997).

3.2.7) Proposed Methods for Registration Evaluation

Both qualitative visual assessment and quantitative landmark analysis were used to evaluate registration accuracy on 20 pre- and post- therapy PET/CT datasets of lymphoma patients. Each of the 20 datasets were grouped into sets of five images, each comprising of a pre-therapy image, an unregistered post-therapy image and three post-therapy images with each of the registrations applied – COM registration, IRTK rigid registration and IRTK non-rigid registration. For each dataset, the pre-therapy image was assigned as image A while the others were randomly assigned as images B, C, D and E to blind observers from the registration method applied on each post-therapy image. PET and CT datasets were treated as different entities so random assignment of images B, C, D and E was different for PET and CT components for each dataset.

For qualitative visual analysis, three researchers with experience of working with PET/CT images analysed the PET and CT images independently from each other and from the other

imaging modality. For each dataset, each researcher ranked images B-E in order of how successfully registered they were with image A and gave each a quality score from 5 (poor) to 1 (excellent) to define the quality of registration according to a defined scoring system (Table 3.1). Rankings were done based on the quality score i.e. the better the quality score, the better the ranking. For images with the same quality scores, ranking was based on the observer's judgement on which was more accurately aligned. The study was completed using HERMES Hybrid Viewer software (Nuclear Diagnostics AB, Stockholm, Sweden) which allowed each observer to view overlay images of each set of pre- and post- therapy scans and make spatially correlated comparisons between the two.

Score	Description
1 (excellent)	Almost perfect fit, no misalignment noticeable, almost like looking at the original baseline image.
2	Very small misalignment is just detectable but generally all the main organs and anatomy are very well aligned.
3	Some misalignment of organs and anatomy is noticeable in certain areas but generally everything is aligned to a suitable standard.
4	Misalignment of organs and anatomy is clear but the body is generally aligned.
5 (poor)	Misalignment of the entire body.

Table 3.1: Quality Score Criteria for Judging Registration Accuracy

CT registration was validated by a quantitative anatomical landmark study by one observer on the same groups of images for the 20 datasets. A number of anatomical landmarks were located on the pre-therapy (image A) and post- therapy (images B-E) images to measure the distance between corresponding points (Table 3.2). The difference in voxels and mm was measured between anatomical landmarks on the pre- therapy image and the same landmark on the post-

therapy image, in each plane, to estimate registration error. The distance between two anatomical locations on a target image and a registered image can be referred to as the target registration error (TRE) and this will be used to estimate the accuracy of registration methods (Fitzpatrick and West, 2001).

When using anatomical landmarks, TRE is affected by observer error in identifying each landmark so the accuracy of the observer for selecting the voxel for each landmark will have to be taken into account. The observer was asked to select the voxel for an anatomical landmark on five pre- and post- therapy images, five different times, to estimate observer error in locating each anatomical landmark. When dealing with fiducial markers, the distance between a given localised point and the actual, real unknown fiducial position of the point is referred to as the fiducial localisation error (FLE) (Fitzpatrick *et al*., 1998b). For the anatomical landmark analysis, the FLE is defined as the error in the observer locating a single fiducial marker, or in this case an anatomical landmark.

Tip of the tail of the spine	End of the trachea
Tip of the left hip/coxal bone	Medial inferior tip of the left lung
Tip of the right hip/coxal bone	Medial inferior tip of the right lung
Central point of the 5th lumbar vertebrae	Lateral inferior tip of the left lung
Central point of the 11th thoracic vertebrae	Lateral inferior tip of the right lung
Central point of the 7th thoracic vertebrae	Superior tip of the liver
Medial point of the left scapula	Inferior tip of the liver
Medial point of the right scapula	Superior tip of the left kidney
Central point of the manubrium	Inferior tip of the left kidney
Tail of the sternum	Superior tip of the right kidney
Superior tip of the left lung	Inferior tip of the right kidney
Superior tip of the right lung	Superior tip of the spleen

Table 3.2: Anatomical Landmarks for Quantitative Registration Accuracy Analysis

3.3) Results of Registering Pre- and Post- Therapy PET/CT Scans

3.3.1) Initial Visual Assessment of Registered Images

A brief visual assessment of overlay and subtraction images gives an idea of the success of registration for each algorithm. It becomes apparent that while the COM method has great success on some individual datasets, it is not robust enough over the 20 datasets to be used usefully as in some cases it is no better aligned than the unregistered image. The IRTK rigid and non-rigid methods show much better success over the 20 datasets, demonstrating more robust algorithms and promising registrations upon visual inspection. The non-rigid algorithm shows more accurate registration when viewing overlay and subtracted images. This is illustrated in Figure 3.1, a subtraction image of dataset 1's pre- and post- therapy CT images, and Figure 3.2, an overlay image of dataset 3's pre- and post- therapy CT images. The COM registration works well on dataset 3 but poorly on dataset 1 where IRTK rigid and non-rigid algorithms appear well aligned on both. It is worth noting the difference in initial misalignment in the transverse plane which is small in dataset 1 (Figure 3.1A) but substantial in dataset 3 (Figure 3.2A).

3.3.2) Anatomical Landmark Analysis Results

3.3.2.1) Fiducial Localisation Error for Anatomical Landmark Analysis

Whole body anatomical landmark analysis, or even registration, is rare in literature. Therefore, with little information or recommendations on what landmarks could be used outside of the brain, a degree of trial and error was used to choose suitable anatomical points (Table 3.2). Typically, it has been shown that between 8-16 landmarks are enough to produce good results for validating images using landmark based techniques (Hill *et al*., 1993; Strasters *et al*., 1997; Hawkes, 1998). However, with the suitability of potential landmarks unknown, a number of hard (skeletal, bones) and soft (organs) landmarks were selected for potential use in this study.

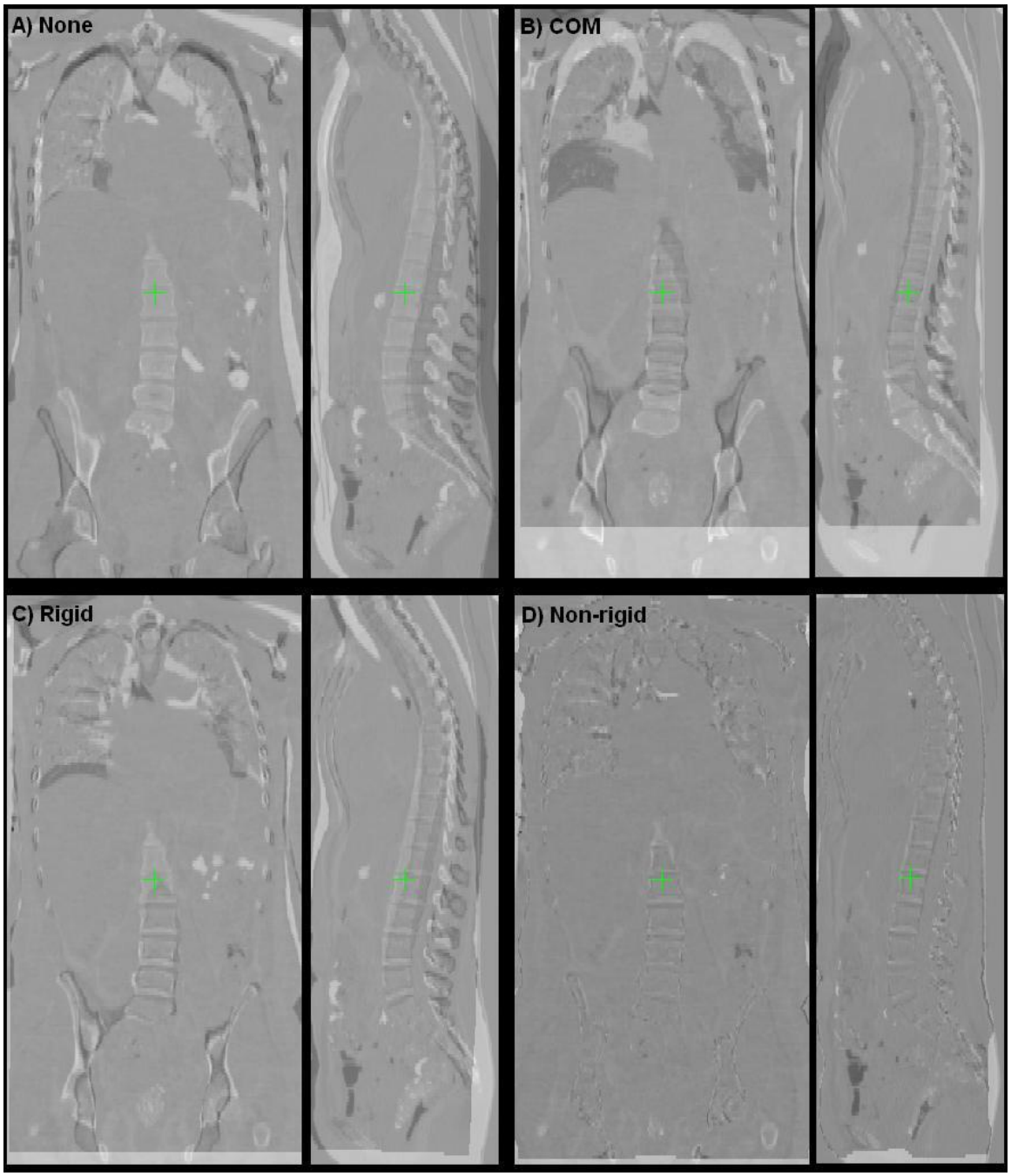


Figure 3.1: Subtraction Images for Registrations on Dataset 1

The four images show the pre-therapy CT scan minus the post therapy CT scan with (A) no-registration, (B) COM registration, (C) rigid registration only, and (D) non-rigid registration applied.

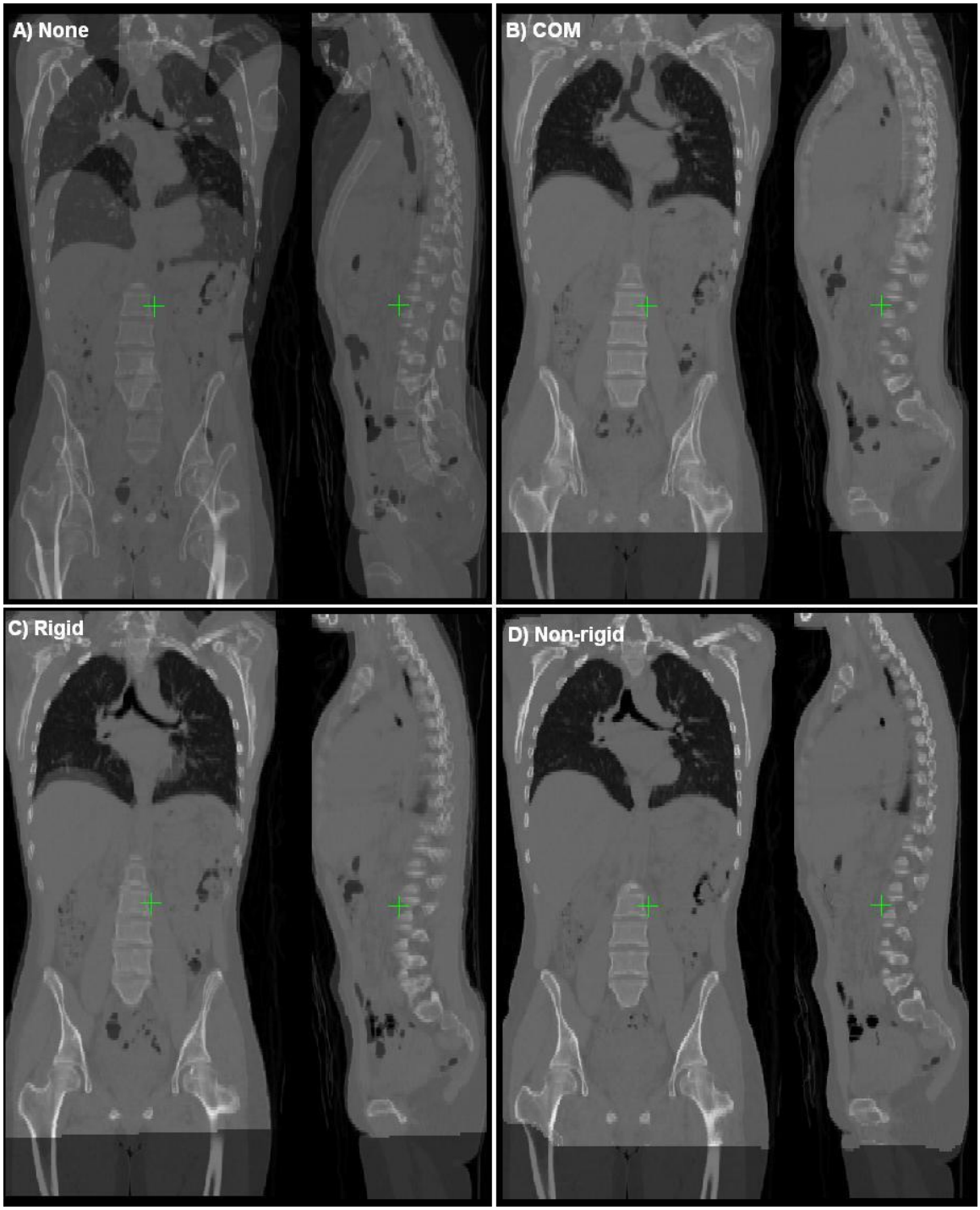


Figure 3.2: Overlay Images for Registrations on Dataset 3

The four images show the pre-therapy CT scan overlayed with the post-therapy CT scan with (A) no-registration, (B) COM registration, (C) rigid registration only, and (D) non-rigid registration applied.

The initial 24 anatomical landmarks (Table 3.2) were located over the first five datasets of pre- and post- therapy CT images, five separate times by the same observer, with a sufficient time gap between each attempt to locate the same landmark. The FLE was defined as the average error in distance between the observer-located voxels. FLE was calculated for each landmark for each 2-D plane and as a 3-D distance incorporating all planes in the image. The mean FLE for each landmark over the five scans is shown in Table 3.3. Landmarks which had a high mean 3-D distance FLE of >5mm were removed from the study, leaving the remaining landmarks suitable for assessing registration accuracy.

Table 3.3 shows six anatomical landmarks proved much harder to reliably identify than the majority of others (landmarks 2, 3, 14, 15, 16 and 17). This could be due to a number of factors including unclear instructions on how to identify the landmark, difficulty in accurately identifying the location of the landmark and variability between scans and patients making a landmark harder to locate. The decision was made to remove these six landmarks as there was a great deal of unreliability in locating them consistently, illustrated by a mean 3-D distance FLE of >5mm. The remaining 18 landmarks were found to have a mean 3-D distance FLE of <5mm and were allocated for use in the study. The mean FLE of the selected 18 landmarks was 1.57mm in the x-plane, 0.93mm in the y-plane and 0.33mm in the z-plane. The mean 3-D distance measurement was just 2.01mm^2. These results show that landmarks can be identified accurately and although there may be some FLE in the results, it should be minimal and typically in the region of just 2mm.

No	Landmark	X	Y	Z	3D Distance	Included
1	Tip of the tail of the spine	2.42	0.90	0.94	2.90	Yes
2	Tip of the left hip/coxal bone	10.74	9.02	10.86	20.11	No
3	Tip of the right hip/coxal bone	11.60	7.03	10.43	19.16	No
4	Central point of the 5th lumbar vertebrae	1.37	0.70	0.20	1.67	Yes
5	Central point of the 11th thoracic vertebrae	1.48	0.62	0.43	1.70	Yes
6	Central point of the 7th thoracic vertebrae	1.25	0.74	0.39	1.59	Yes
7	Superior tip of the left clavicle	0.43	0.08	0.16	0.51	Yes
8	Superior tip of the right clavicle	0.31	0.00	0.12	0.43	Yes
9	Central point of the manubrium	2.58	1.33	0.70	3.21	Yes
10	Tail of the sternum	1.25	0.86	0.59	1.89	Yes
11	End of the Trachea	0.51	0.47	0.27	0.90	Yes
12	Superior tip of the left lung	1.13	1.56	0.12	2.10	Yes
13	Superior tip of the right lung	0.86	0.86	0.00	1.28	Yes
14	Medial inferior tip of the left lung	13.24	6.80	2.70	15.44	No
15	Medial inferior tip of the right lung	11.99	6.87	2.97	14.76	No
16	Lateral inferior tip of the left lung	15.86	10.94	0.59	20.18	No
17	Lateral inferior tip of the right lung	27.07	11.13	1.33	29.81	No
18	Superior tip of the liver	0.94	0.55	0.00	1.15	Yes
19	Inferior tip of the liver	3.91	2.19	0.51	4.86	Yes
20	Superior tip of the left kidney	2.30	1.13	0.08	2.62	Yes
21	Inferior tip of the left kidney	1.87	1.17	0.35	2.35	Yes
22	Superior tip of the right kidney	1.84	1.41	0.47	2.45	Yes
23	Inferior tip of the right kidney	1.37	0.78	0.12	1.64	Yes
24	Superior tip of the spleen	2.42	1.37	0.59	2.93	Yes
MEAN (ALL)		4.95	1.45	2.85	6.49	
S.D. (ALL)		6.67	2.93	3.57	8.36	
MEAN (18 LANDMARKS SELECTED)		1.57	0.93	0.33	2.01	
S.D. (18 LANDMARKS SELECTED)		0.91	0.26	0.53	1.08	

Table 3.3: Fiducial Localisation Error for Anatomical Landmarks

Mean fiducial localisation error (FLE) in X, Y and Z planes and as a 3-D distance. Values are in mm worked out from voxel locations. Light grey values indicate those <5mm, grey between 5-10mm and dark grey those >10 mm. 3-D distances are in mm. All landmarks were used every time in each dataset.

3.3.2.2) Target Registration Error for Anatomical Landmark Analysis

The 18 anatomical landmarks were identified on all 100 CT images of pre-therapy scans, unregistered post-therapy scans and registered post-therapy scans using all three registration methods. Occasionally, some landmarks could not be identified well enough to be deemed suitable for analysis in either one or more of the images. This was because they were considered to be too difficult to identify accurately or beyond the boundaries of the subvolume. In total, 301 landmarks were used for assessing registration accuracy. For each dataset, there was at least 11/18 landmarks selected (median: 15 landmarks). The difference in distance between the anatomical landmark on the pre-therapy CT image and on each of the four post-therapy CT images was calculated, in each direction and as a 3-D distance i.e. a combination of the difference in distance in all three planes (x, y and z planes), to provide an estimate of the TRE to assess the accuracy of registration. The TRE on each plane was taken as the distance from the voxel selected by an observer on the target image to represent a given landmark to the same voxel located on the registered image to represent the same landmark,. 3-D distance TRE was calculated as:

$$TRE = \frac{\sum_{i=1}^{n} i \sqrt{x_i^2 + y_i^2 + z_i^2}}{n} \qquad [3.2]$$

where x_i, y_i and z_i are the difference in distance between the landmark in the pre-therapy and post-therapy CT images, in mm, summed over n number of subjects.

The mean TRE, in each plane and as a 3-D distance, for each anatomical landmark over the 20 datasets for all registration methods is presented in Table 3.4. The mean TRE, in each direction and as a 3-D distance, for each dataset over the 18 anatomical landmarks for all registration methods is presented in Table 3.5. TREs are calculated for each of the four categories of

registered images: (i) no registration/unregistered, (ii) COM registration, (iii) IRTK rigid registration and (iv) IRTK non-rigid registration. The results show that prior to any image registration, the mean TRE between pre- and post- therapy CT scans was ~40mm (using 3-D distance measure). The non-rigid registration algorithm was the most accurate at registering CT images with a mean TRE of ~6.5mm between pre-therapy and non-rigidly registered post-therapy CT scans. Rigid registration also showed a reasonable registration accuracy of ~10mm.

Unregistered post-therapy images had a TRE of >8mm in all planes with particular misalignment in the transverse plane, with a mean TRE of ~33mm. Although COM registration has been shown to be very favourable on some occasions, it has also produced further misalignment compared to unregistered images on others. On average, COM proved to be only slightly better than no registration. It can be an efficient way of performing registration, but its lack of robustness means it is not a reliable option. Subvolume selection and field of view is potentially an issue with the COM method as pre- and post- therapy scans can show different areas of anatomy. There may be ways to improve this method for future use but the rigid registration method performs much better and is not much more computationally exhaustive. While registration as accurate as 1-2mm is common in the brain, patient movement and natural deformation can cause more issues when registering whole body images and so to accurately align images within 5-10mm, comparable with PET resolution, is a promising result.

The anatomical landmark analysis shows variations in TRE when using different landmarks for rigid and non-rigid registration. Observer error is inevitable going to affect this but it's worth noting landmarks such as the tip of the clavicles, tail of the spine and tips of the kidneys have larger TREs than landmarks on the vertebrae and in the lung. Differences in TRE between different landmarks on the same image may be because of differences in the registration accuracy in different regions of the body (TRE) or the ability of the observer to successfully

locate landmarks (FLE). However, with relatively low FLEs found between identifying anatomical landmarks (Table 3.3), it is likely these differences are primarily down to TREs, where some landmarks are not registered as well as others. For example, the tips of the clavicles and kidneys have much higher mean TREs than the three vertebrae, however, particularly for rigid registration, the tips of the clavicles and kidneys are less likely to be aligned in registered images in comparison to the centrally positioned (in comparison to the clavicles) and rigid (in comparison to the kidneys) vertebrae. Registration accuracy is fairly consistent throughout the 20 datasets with only a few datasets with mean TRE 3-D distances >10mm. IRTK rigid and non-rigid registration algorithms improved alignment in all 20 datasets with excellent improvement in the majority of datasets. To improve registration even further, a smaller subvolume of the image could be used when analysing a specific tumour, or a region of several tumours, as this reduces the possibility of misregistrations due to registration algorithms registering certain anatomy in the subvolume better than the actual area of interest.

			NO REGISTRATION				COM REGISTRATION				IRTK RIGID REGISTRATION				IRTK NON-RIGID REGISTRATION			
#	Landmark	No	X	Y	Z	3-D	X	Y	Z	3-D	X	Y	Z	3-D	X	Y	Z	3-D
1	Tail of the spine	12	4.88	11.31	30.79	37.31	5.62	6.51	14.99	19.37	3.17	2.60	3.27	6.40	2.69	2.36	6.27	8.29
2	5th lumbar vertebrae	20	7.52	9.52	32.37	38.24	5.13	5.81	30.90	34.13	2.88	2.54	5.07	7.22	1.37	0.93	2.29	3.39
3	11th thoracic vertebrae	20	5.91	11.33	33.03	39.13	4.49	6.15	31.07	34.62	4.25	2.10	4.74	8.06	1.37	1.07	4.58	5.65
4	7th thoracic vertebrae	20	6.79	11.43	32.05	38.01	6.79	7.67	30.25	34.63	2.59	1.46	4.58	6.30	1.81	0.88	2.78	4.52
5	Superior tip of the left clavicle	11	13.32	11.81	44.29	51.22	13.67	7.19	39.83	46.87	5.50	6.21	7.73	13.70	4.88	5.15	8.62	12.89
6	Superior tip of the right clavicle	11	12.52	16.87	42.81	51.79	12.34	8.79	40.73	48.09	6.04	5.15	11.00	14.84	5.33	4.26	8.03	11.59
7	Central point of the manubrium	20	8.98	9.23	34.01	38.65	10.74	6.69	29.92	35.64	3.56	2.44	5.07	7.27	2.39	1.46	2.29	4.36
8	Tail of the sternum	17	9.59	6.95	32.32	37.22	6.89	4.31	21.16	24.66	6.03	2.76	4.23	8.94	2.70	2.41	4.04	6.28
9	End of the Trachea	20	7.86	9.86	31.72	36.94	9.62	6.74	31.07	35.43	3.66	2.00	7.03	8.87	1.12	1.03	1.80	2.94
10	Superior tip of the left lung	17	11.55	10.80	31.55	40.35	10.23	8.33	27.31	35.93	5.11	3.04	5.00	8.81	2.64	1.95	0.38	3.95
11	Superior tip of the right lung	17	9.71	15.97	32.51	42.94	9.48	7.98	30.20	35.51	3.68	3.79	5.19	8.57	2.07	1.55	0.77	3.05
12	Superior tip of the liver	19	8.89	11.00	30.63	38.45	7.35	7.25	24.78	29.37	7.61	5.50	7.06	13.45	3.44	2.67	0.69	4.97
13	Inferior tip of the liver	9	13.02	9.55	31.61	39.68	13.56	10.09	27.25	35.57	7.27	6.84	8.36	15.48	7.81	2.28	7.27	11.75
14	Superior tip of the left kidney	19	7.14	11.51	31.15	37.81	5.24	6.53	30.63	34.23	5.81	3.70	7.23	11.04	4.01	2.57	6.20	8.55
15	Inferior tip of the left kidney	19	9.87	11.98	31.67	39.38	6.42	7.04	29.77	33.88	5.50	4.06	9.12	12.87	6.12	3.44	7.23	11.35
16	Superior tip of the right kidney	20	7.28	11.04	35.48	42.31	6.79	6.64	31.23	35.55	5.47	3.76	6.21	10.14	3.91	3.08	6.21	9.05
17	Inferior tip of the right kidney	20	8.94	7.76	37.44	43.13	7.18	5.57	34.17	37.35	4.69	4.44	8.18	11.50	4.20	3.52	5.72	8.75
18	Superior tip of the spleen	10	10.35	9.67	20.60	29.40	10.55	17.87	14.72	31.57	5.37	2.64	6.21	11.31	3.03	2.93	0.65	4.74
ALL		301	8.79	10.85	33.07	39.89	8.00	7.23	29.19	34.37	4.78	3.44	6.28	9.93	3.16	2.28	3.99	6.62

Table 3.4: Target Registration Error for Anatomical Landmarks

Mean target registration error (TRE) over the 20 datasets for the 18 anatomical landmarks for each registration method in X, Y and Z planes and as a 3-D distance. Values are in mm worked out from voxel locations. Light grey indicate those <5mm, grey between 5-10mm and dark grey are those >10 mm. 3-D distance are in mm. # = the number of the landmark used and No = the number of times that landmark was used in the analysis of the 20 datasets. Values in the ALL row are mean values apart from the number of times a landmark was used which is a total.

		NO REGISTRATION				COM REGISTRATION				IRTK RIGID REGISTRATION				IRTK NON-RIGID REGISTRATION			
Dataset	**Landmarks Used**	**X**	**Y**	**Z**	**3-D**	**X**	**Y**	**Z**	**3-D**	**X**	**Y**	**Z**	**3-D**	**X**	**Y**	**Z**	**3-D**
1	11	23.17	3.55	9.81	27.80	9.14	4.08	48.46	50.42	9.77	2.93	9.51	14.67	5.95	1.95	6.84	10.00
2	17	5.11	3.62	6.73	10.32	5.11	2.87	9.62	12.21	3.96	2.99	2.50	6.90	3.04	3.04	2.50	5.79
3	17	9.88	18.32	75.98	79.58	8.16	6.55	5.58	13.24	5.28	3.68	4.62	9.22	2.59	1.67	2.31	4.59
4	17	13.50	10.17	12.50	22.86	13.90	12.93	17.89	32.16	9.71	5.86	16.35	22.02	6.84	3.33	12.70	16.39
5	16	10.93	6.41	22.69	26.52	4.52	3.72	10.01	12.56	3.17	2.87	4.29	6.95	2.50	1.59	2.45	4.56
6	16	11.66	18.13	5.31	23.35	4.64	5.55	17.99	20.08	3.85	3.11	5.93	8.97	2.81	1.65	2.86	5.42
7	12	2.20	10.50	48.51	50.13	8.79	8.30	29.98	33.02	2.12	2.44	3.54	5.49	1.63	1.87	5.72	6.99
8	14	3.77	11.58	78.25	79.75	2.44	8.30	97.63	98.18	3.63	4.81	5.61	9.20	3.14	1.95	1.87	5.05
9	13	11.49	16.30	19.62	30.24	8.11	40.26	49.05	64.78	3.83	6.46	6.79	11.41	2.48	2.70	3.02	5.54
10	15	8.98	6.12	38.15	40.48	9.96	7.42	13.52	20.55	7.88	3.52	7.85	13.16	3.71	2.08	5.67	7.96
11	14	8.44	8.51	79.88	80.89	2.72	2.51	92.96	93.10	2.79	2.23	7.01	8.51	1.74	2.58	5.14	6.91
12	17	7.58	18.32	15.20	26.14	11.09	4.88	6.54	14.89	4.83	2.53	5.58	8.58	2.59	1.67	2.89	4.67
13	14	8.09	12.49	67.97	72.21	14.86	3.07	11.91	20.97	5.16	2.79	4.67	8.17	3.56	1.60	2.80	5.43
14	17	8.33	8.67	6.16	15.57	8.16	5.51	6.16	13.14	4.37	2.47	5.96	8.88	3.45	2.41	2.69	6.10
15	14	20.65	7.18	45.55	52.48	20.72	5.30	110.5	113.4	10.25	4.95	8.18	15.84	6.77	4.12	6.77	11.72
16	14	5.51	2.93	6.31	9.91	7.11	2.37	35.27	36.28	4.53	2.23	2.34	6.47	2.09	1.46	1.87	4.00
17	17	5.74	21.71	15.77	28.56	7.98	7.01	10.19	16.23	2.36	3.39	7.12	9.05	1.67	2.41	2.89	4.88
18	17	2.76	2.36	62.52	62.69	3.45	2.41	10.77	12.24	1.15	1.95	3.65	5.38	1.15	1.26	2.31	3.63
19	15	6.32	15.30	46.22	50.07	5.08	7.88	33.14	35.30	4.30	4.43	6.54	10.42	3.26	3.19	4.36	7.66
20	14	5.30	11.65	6.54	16.59	5.16	7.88	7.01	13.24	3.77	3.35	7.71	9.91	2.93	3.14	3.04	6.20
ALL	301	8.79	10.85	33.07	39.89	8.00	7.23	29.19	34.37	4.78	3.44	6.28	9.93	3.16	2.28	3.99	6.62

Table 3.5: Target Registration Error for Each Dataset

Mean target registration error (TRE) over the 18 anatomical landmarks for each of the 20 datasets for each registration method in X, Y and Z planes and as a 3-D distance. Values are in mm worked out from voxel locations. Light grey indicate those <5mm, grey between 5-10mm and dark grey are those >10 mm. 3-D distance are in mm. Landmarks Used = the number landmarks used in the analysis of that dataset. Values in the ALL row are mean values apart from the number of times a landmark was used which is a total.

3.3.3) Visual Assessment Results

Results of the visual analysis study completed by three observers ranked the non-rigid registration method the highest 98% of the time on CT images and 82% of time on PET images. For both PET and CT, in cases where the non-rigid method was not ranked first, the rigid registration method was. The rigid and non-rigid registrations consistently got high quality scores with mean scores of 2.80 (CT) and 3.25 (PET) for rigid and 1.77 (CT) and 2.10 (PET) for non-rigid (Table 3.6). Images with no registration had poor quality scores of 4.67 (CT) and 4.53 (PET), as did COM registration which performed only marginally better than no registration with average quality scores of 4.32 (CT) and 3.83 (PET). Registration accuracy was consistent over the 20 patients with mean observer quality scores for each image between 1.34-2.67 (PET) and 1.34-3 (CT) for non-rigidly registered images and from 2.34-3.34 (PET) and 2.34-4.34 (CT) for rigidly registered images. No registration saw a range of 3.34-5 (PET and CT) while COM registered images showed the most variability ranging from 2.34-5 (PET) and 3.34-5 (CT).

Fleiss' Kappa analysis, a statistic for measuring the agreement of a number of observers when giving categorical ratings, was used to measure observer variability between ranking and quality scores (Fleiss, 1971; Landis and Koch, 1977). PET and CT rankings have substantial agreement between observers (>0.600), while PET and CT quality scores have a fair agreement (>0.200) according to suggested criteria based on the opinions of the authors of the methodology (Landis and Koch, 1977), however, it should be noted that this criteria is purely based on opion and is not universally accepted. The effects of registration and the observer were explored using linear fixed-effects modelling to fit effects to data sampled from normal distributions through the MIXED procedure in SPSS Statistics Version 20 (IBM, Armonk, NY, USA). Registration was treated as a fixed effect while the observer was treated as a random effect. Both effects of registration and registration and observer interaction were found to be significant, $F_{3,152.6}$ =

165.799 ($p < 0.001$) and $F_{8,88.8} = 4.3$ ($p < 0.001$) respectively. Non-rigid registrations were consistently given quality scores of 1-3 for 97% of the scans on PET and 98% on CT, showing good consistency and robustness of registration. However, the score they were given seemed to change depending on the observer, potentially down to each observer's interpretation of the quality score criteria. Observers performed better when ranking images as shown by a substantial agreement in Fleiss' Kappa analysis.

METHOD	PET RANKING				PET QUALITY SCORE			
	MEAN	MED	MODE	S.D.	MEAN	MED	MODE	S.D.
NO REGISTRATION	3.73	4	4	0.48	4.53	5	5	0.68
CENTRE OF MASS	3.13	3	3	0.57	3.83	4	4	0.99
RIGID	1.95	2	2	0.59	2.80	3	3	0.58
NON-RIGID	1.18	1	1	0.39	2.10	2	2	0.75
ALL	N/A	N/A	N/A	1.12	3.32	3	3	1.21
	FK = 0.611 (0.538 - 0.684)				FK = 0.367 (0.298 - 0.436)			
METHOD	**CT RANKING**				**CT QUALITY SCORE**			
	MEAN	MED	MODE	S.D.	MEAN	MED	MODE	S.D.
NO REGISTRATION	3.80	4	4	0.44	4.67	5	5	0.63
CENTRE OF MASS	3.13	3	3	0.43	4.32	4	4	0.75
RIGID	2.05	2	2	0.34	3.25	3	3	0.82
NON-RIGID	1.02	1	1	0.13	1.77	2	2	0.77
ALL	N/A	N/A	N/A	1.12	3.50	4	5	1.35
	FK = 0.850 (0.777 - 0.923)				FK = 0.280 (0.213 - 0.347)			

Table 3.6: Visual Analysis Results for CT and PET Scans for All Registration Methods

Results are mean values of all three observers over all 20 datasets. Ranking is scored from 1 (best) to 4 (worst) and quality scores are rated from 1 (excellent) to 5 (poor). MED = Median, S.D. = Standard Deviation, FK = Fleiss' Kappa (with 95% confidence intervals).

3.3.4) PET/CT Response Criteria Results

Successful registration was achieved in patients with a variety of responses on both PET/CT and CT (Table 3.7). Response was judged by using both anatomical and metabolic response criteria, as defined by the IWC on CT and Deauville criteria on PET (Cheson *et al*., 1999; Meignan *et al*., 2009). Table 3.7 shows response assessment for both PET and CT criteria for each of the 20 datasets. The degree of change in tumour size can potentially have an effect on the success of the registration of an image. Larger changes in the patient from one scan to another presents more of a challenge for the registration, and in terms of non-rigid registration can cause unwanted deformations in order to match changes in tumour size and shape. Changes in tumour size in this dataset can be quite drastic with CT measurements decreasing by up to 30mm in size (~70%). On average, IWC measurements were found to approximately half from a pre-therapy average of 29mm to a post-therapy average of 13mm for each lesion. These changes are more drastic than anticipated. Despite the changes in metabolic activity on PET and nodal size on CT, registration comparable with PET spatial resolution can be achieved using the non-rigid registration method.

It was hoped that changes of tumour size on CT would be small allowing accurate registration to take place. While changes on CT are greater than expected, changes on PET are still more drastic in comparison. To illustrate this, Figure 3.3 shows two datasets of corresponding PET/CT pre- and post- therapy images with different responses. For both datasets, there is reduction in measured tumour size on CT between pre- and post- therapy images, particularly on dataset 2. However, on PET, the reduction of uptake is greater, with no uptake on post-therapy scans.

Dataset and Patient Information				Deauville (PET)		IWC (CT)				
No	Age	Sex	Weight (kg)	Score	Response	Node	Pre	Post	Difference (in %)	Response
1	27	F	63	2	Negative	1	44	13	31 (71)	CRu
2	55	M	82	1	Negative	2 3	18.3 16.3	5 5	13.3 (73) 11.3 (69)	CR
3	21	F	51	1	Negative	4	16.3	11	5.3 (33)	PR
4	21	M	80	1	Negative	5 6 7	50 34 19	15 7 7	35 (70) 27 (79) 12 (63)	PR
5	43	F	59	5	Positive	8	51	31	20 (39)	PR
6	39	F	83	4	Positive	9	70	46	24 (34)	PR
7	27	M	57	1	Negative	10 11 12	15 20 17	5 5 5	10 (67) 15 (75) 12 (71)	CR
8	19	M	76	1	Negative	13 14	15 40	7 21	8 (53) 19 (48)	PR
9	66	F	57	1	Negative	15 16 17	18 17 16	6 5 5	12 (67) 12 (71) 11 (69)	CR
10	63	F	82	4	Positive	18 19	33 34	17 9	16 (49) 25 (74)	PR
11	31	F	55	2	Negative	20 21 22	29 37 19	13 27 10	16 (55) 10 (27) 9 (47)	SD
12	28	M	77	1	Negative	23 24	20 17	8 5	12 (60) 12 (71)	CR
13	18	F	62	5	Positive	25	82	53	29 (35)	PR
14	49	M	90	1	Negative	26 27	35 23	5 8	30 (86) 15 (65)	CR
15	21	M	55	5	Positive	28 29	27 15	5 4	22 (82) 11 (73)	CR
16	25	M	85	4	Positive	N/A	N/A	N/A	N/A	NE
17	56	F	70	1	Negative	30	18	11	7 (39)	SD
18	23	M	68	1	Negative	31 32	23 19	13 7	10 (44) 12 (63)	PR
19	51	M	82	1	Negative	33 34 35 36	37 52 39 29	16 28 27 22	21 (57) 24 (46) 12 (31) 7 (24)	SD
20	70	M	65	5	Positive	37 38	22 18	9 6	13 (59) 12 (67)	CR

Table 3.7: Response Assessment Results using Deauville Criteria (PET) and IWC (CT)

The response for each dataset is listed according to the Deauville score and the IWC criteria. For Deauville criteria, a score of 1-3 is regarded as 'negative' and 4-5 as 'positive' for lymphoma. For IWC Criteria, CR = Complete Response, CRu = Complete Response (unconfirmed), PR = Partial Response, SD = Stable Disease and NE = Non-Evaluable. Nodal measurements for IWC are of the longest transverse diameter of the node in mm.

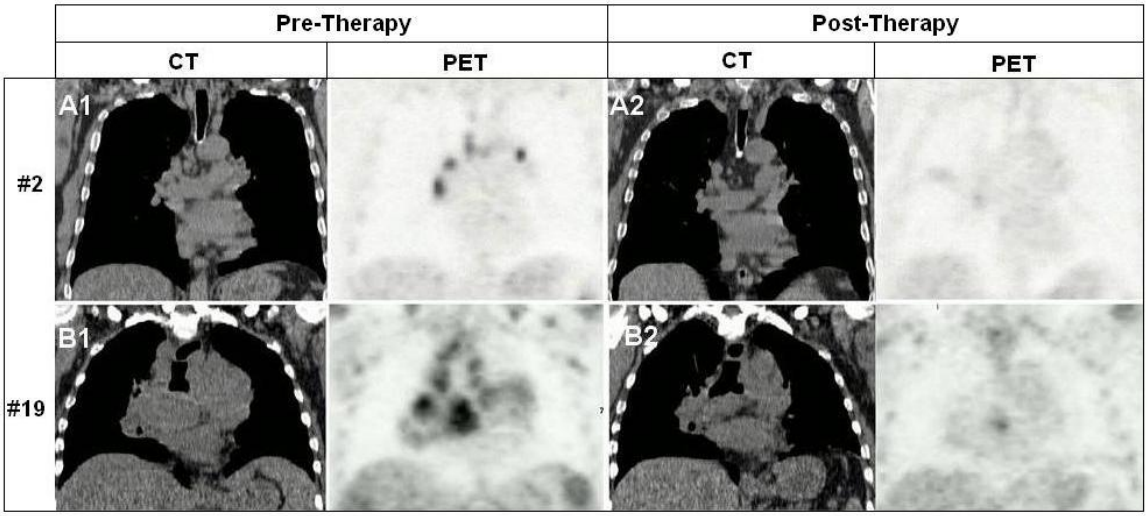


Figure 3.3: Visual Comparison of Change between PET and CT

The images show (A1) pre-therapy PET/CT and (A2) post-therapy PET/CT for a dataset for which IWC on CT indicates complete response as does visual inspection on PET (Dataset 2) and (B1) pre-therapy PET/CT and (B2) post-therapy PET/CT for a dataset for which IWC on CT indicates stable disease but visual inspection on PET shows response (Dataset 19).

3.3.5) Summary of Findings on Registration Accuracy

Both the qualitative visual assessment and quantitative anatomical landmark results show that the non-rigid registration algorithm provides the greatest improvement in alignment of CT and PET images. The landmark analysis results on CT and good registration quality scores suggest that PET images are aligned with accuracy comparable to their resolution and therefore have the potential to provide useful information in identifying response to therapy. It was considered that initial patient positioning could affect the success of registrations. However, the quantitative study shows almost no correlation between no-registration and rigid (Pearson correlation coefficient (pcc) = 0.009, for TRE 3-D distances over all landmarks on all 20 scans) or non-rigid (pcc = 0.0324) registration. Therefore, the initial positioning of the patient has no bearing on

success of registration. This study has shown that misalignment of pre- and post- therapy PET scans can frequently be ~40mm. Improved PET-PET image alignment achieved using image registration could help aid clinicians with visual analysis, either by helping them find corresponding tumours more easily or by giving them extra information in terms of the areas of response.

There is potential for more work on registration algorithms. For example, the use of CT transformations being applied to PET images relies on pre- and post- therapy CT images being accurately aligned with pre- and post- therapy PET images. Visual registration results are positive for both PET and CT, and studies showing that there are usually only marginal misalignments between concurrent PET/CT scans. However, there is still potential for some mismatch, particularly in certain cases where there are significant differences in respiration between scans which would be especially relevant when studying lung tumours for example. Therefore, further work on adding a PET to CT registration stage into the rigid and non-rigid IRTK algorithms would be beneficial.

Additional testing of the rigid and non-rigid registrations on different datasets of differing diseases should be conducted to ensure the algorithm's flexibility. Although the methods are designed to be robust, this needs to be validated fully. Testing of the algorithms using finer resolutions, more iterations and cubic rather than linear interpolation could improve the registration accuracy. However, these changes would most likely severely affect efficiency and the time taken to register images, although, it could be useful when looking at smaller volumes of specific areas of disease. It may also be beneficial to modify the non-rigid algorithm so that hard structures, such as the skeleton, are kept as rigid transformations as there should be little deformation in these areas (Little *et al*., 1997).

3.4) Affect of Registration of PET/CT scans on SUVs

3.4.1) Importance of SUV Changes on Registered PET Images

Registration of PET/CT studies is likely to be of benefit when it comes to identifying response to therapy as looking at registered images can allow the viewer to see both pre- and post- therapy scans in the same anatomical space, making it easier to compare lesions from one scan to the next. There is also the potential to use registered PET images to quantify the change between pre- and post- therapy uptake in the same tumour area using techniques such as voxel-by-voxel analysis. However, registration transformations can cause changes in SUVs and the shape of PET structures on the registered image compared to the unregistered image. Ideally, SUVs and tumour size should be consistent between registered and unregistered post-therapy images for a fair comparison of analysis between pre- to post- therapy images with just the tumour location changing so it is aligned between the two. Therefore, it is important to know the scope of the changes in SUV and volume of PET structures between pre- and post- therapy images to know how much they could affect user interpretation of a registered image and, more importantly, the potential impact on quantitative techniques. For example, an unregistered post-therapy image may have a SUV_{max} of one value but when a registration transformation is applied it may change and this could have an affect when being interpreted by a clinician or when using quantitative analysis. When using novel registration analysis methods to identifying response to therapy, it is important that registered post-therapy images have not had major, unwarranted changes in SUVs or TVs because of registration rather than physiological changes due to therapy.

There has been very little research on the effect of registration on PET SUVs, particularly when registering pre- and post- therapy PET/CT scans. The difference between rigid and non-rigid registration for aligning PET to CT in patients with lung cancer was investigated and there were no significant differences found in SUV_{max} between algorithms but there was potential for non-

rigid algorithms to cause changes of up to 3ml in TV depending on the breathing procedure used in CT acquisition (Grgic *et al*., 2011). The study concluded that non-rigidly registered post-therapy PET scans achieved using the CT component of a PET/CT scan could not be used reliably to provide unchanged TV measures between unregistered and registered post-therapy images when the CT scan was performed during inspiration. TVs were segmented using a S/B algorithm using scanner parameters, background-SUV and a 70% isocontour of SUV_{mean}. The image registration used was performed in HERMES software using a rigid linear algorithm, with NMI as a similarity measure, and a non-rigid nonlinear warping algorithm, using thin-plate splines (Grgic *et al*., 2009). Another study investigating various registration algorithms on PET/CT test-retest scans of 11 patients with colorectal carcinoma found there was no significant difference in SUV_{max}, SUV_{mean}, TV or TLG between registered and unregistered scans (van Velden *et al*., 2012). However, neither of these studies investigated clinical pre- and post- therapy PET/CT studies and the possible effect of registration transformations.

3.4.2) Methodology for Evaluating SUV and Volume Changes on PET Images

For each of the 20 datasets of patients with lymphoma, the unregistered post-therapy images were compared with IRTK rigidly and non-rigidly registered post-therapy images to assess the changes in SUV_{max}, SUV_{peak} and volume for tumours and other areas of physiological uptake in the image. While taking data from other PET structures, typically the heart, kidneys and bladder, is a useful indication of changes due to registration, differences in tumours are of more relevance and importance as these are areas of interest when assessing response to therapy. Using registered post-therapy images is not ideal, as due to successful treatment there are only a small number of tumours present in post-therapy images over the 20 datasets. Therefore, to obtain more data regarding potential changes in tumour parameters, registration algorithms were run in reverse i.e. registering pre-therapy images to post-therapy images, using the same IRTK registration techniques to ensure more data on the changes in tumours could be obtained. The data analysis

was completed in PETTRA software, described in chapter 2. SUV_{max} was obtained for each PET structure as the maximum value in the TV. TV was delineated using the fixed 2.5 SUV region segmentation method in PETTRA. For this study, SUV_{peak} was defined as the SUV_{max} and its 26-connected neighbours.

3.4.3) Results of SUV and Volume Changes on PET Images

3.4.3.1) Results of SUV and Volume Changes on Post-Therapy Registered Images

For each of the 20 datasets, unregistered post-therapy images were compared with IRTK rigidly and non-rigidly registered post-therapy images to quantify changes in SUV_{max}, SUV_{peak} and TV. The results are shown in Table 3.8 where mean values over the 20 datasets and absolute mean differences and percentage differences were calculated. There were a total of 47 distinguishable areas of uptake. However, only five of these were tumours. IRTK rigidly registered post-therapy images showed no significant change in values from the unregistered post-therapy images with differences across all parameters and all uptake areas being, on average, <2% with the difference always <0.5 SUV. For SUV_{max}, there was no change in any of the tumour values while for SUV_{peak} only two tumours had differences of 1% and 5% with absolute changes <0.2 SUV. TV differences were always <1%. For IRTK non-rigidly registered images, there were only minor changes in SUV_{max} values, particularly in tumours where there was a difference in only one tumour of just 0.8%. SUV_{peak} values had a mean difference of ~5% in tumours, however, absolute differences were <0.3 SUV. Differences in TV were more significant with changes in all five tumours, two of which showed differences of 7.3ml (35%) and 16.2ml (87%).

Uptake Area	No	IRTK RIGID REGISTRATION				IRTK NON-RIGID REGISTRATION			
		SUV_{max}				SUV_{max}			
		Unreg	RR	Abs Diff	Abs Diff %	Unreg	NR	Abs Diff	Abs Diff %
Tumour	5	8.14 (4.31)	8.14 (4.31)	0.00 (0.00)	0.00 (0.00)	8.14 (4.31)	8.14 (4.31)	0.01 (0.02)	0.16 (0.35)
Bladder	14	26.79 (13.63)	26.81 (13.63)	0.02 (0.06)	0.08 (0.31)	26.79 (13.63)	26.72 (13.54)	0.11 (0.34)	0.31 (0.90)
Kidney	16	35.54 (28.61)	35.54 (28.61)	0.00 (0.00)	0.00 (0.00)	35.54 (28.61)	34.73 (27.70)	0.82 (3.27)	1.27 (5.07)
Heart	12	11.35 (4.46)	11.37 (4.47)	0.03 (0.09)	0.20 (0.58)	11.35 (4.46)	11.35 (4.49)	0.05 (0.10)	0.45 (0.90)
Total	47	23.84 (21.06)	23.85 (21.06)	0.01 (0.06)	0.08 (0.34)	23.84 (21.06)	23.54 (20.49)	0.32 (1.91)	0.66 (3.00)
Uptake Area	No	SUV_{peak} (27 voxels)				SUV_{peak} (27 voxels)			
		Unreg	RR	Abs Diff	Abs Diff %	Unreg	NR	Abs Diff	Abs Diff %
Tumour	5	4.59 (1.69)	4.54 (1.69)	0.05 (0.09)	1.13 (2.04)	4.59 (1.69)	4.51 (1.84)	0.17 (0.12)	5.18 (5.48)
Bladder	14	18.29 (8.75)	18.34 (8.72)	0.06 (0.21)	0.40 (1.50)	18.29 (8.75)	18.90 (9.80)	0.97 (1.88)	4.39 (7.31)
Kidney	16	18.52 (15.80)	18.41 (15.83)	0.19 (0.34)	0.82 (2.30)	18.52 (15.80)	17.54 (15.41)	1.02 (2.89)	6.37 (11.25)
Heart	12	7.81 (2.88)	7.90 (3.00)	0.12 (0.23)	1.21 (2.04)	7.84 (2.88)	7.42 (2.98)	0.51 (0.50)	7.58 (8.11)
Total	47	14.24 (11.73)	14.23 (11.72)	0.09 (0.26)	0.83 (1.96)	14.24 (11.73)	13.97 (11.82)	0.78 (1.97)	5.96 (8.73)
Uptake Area	No	Volume (Fixed 2.5 SUV Threshold Segmentation)				Volume (Fixed 2.5 SUV Threshold Segmentation)			
		Unreg	RR	Abs Diff	Abs Diff %	Unreg	NR	Abs Diff	Abs Diff %
Tumour	5	20.84 (24.94)	20.81 (24.82)	0.09 (0.12)	0.29 (0.29)	20.84 (24.94)	23.11 (25.51)	5.29 (6.75)	30.38 (33.89)
Bladder	14	118.8 (94.44)	118.6 (94.67)	0.41 (1.11)	0.50 (1.56)	118.9 (94.44)	123.8 (101.9)	10.85 (10.15)	10.27 (8.60)
Kidney	16	59.10 (27.93)	59.14 (27.88)	0.13 (0.19)	0.27 (0.39)	59.10 (27.93)	60.93 (26.94)	5.04 (5.54)	10.47 (11.79)
Heart	12	207.6 (94.40)	207.4 (94.32)	0.29 (0.18)	0.17 (0.14)	207.6 (94.40)	211.6 (104.4)	24.96 (24.95)	10.56 (7.71)
Total	47	110.7 (96.08)	110.6 (96.07)	0.25 (0.62)	0.32 (0.88)	110.7 (96.08)	114.1 (101.1)	11.88 (16.06)	12.55 (14.80)

Table 3.8: Changes in SUV_{max}, SUV_{peak} and Volume between Unregistered Post-Therapy Scans and IRTK Registered Post-Therapy Scans

All parameters are given as mean values for each uptake area over the 20 datasets. SUV_{max} is the intensity of the maximum voxel in the uptake area, SUV_{peak} represents the mean of the SUV_{max} and its 26-connected neighbouring voxels, and the volume (in ml) is segmented using fixed 2.5 SUV threshold region growing. Unreg = Unregistered post-therapy values, RR = rigidly registered post-therapy values, NR = non-rigidly registered post-therapy values, Abs Diff = absolute difference between unregistered and registered values, Abs Diff % = % difference between unregistered and registered values. (S.D. in brackets).

Further investigation into these tumours shows why there are large changes in TV between unregistered and registered post-therapy images. The difference of 7.3ml (35%) in TV in dataset 3 can be explained by the deformation in the non-rigidly registered image creating a gap in the connectivity of the segmentation. This has led to some of the disease not being segmented in the non-rigidly registered image. With another segmentation added to the existing segmentation to take this into account, the difference in TV is 0.9ml (4%). The change in TV in dataset 5 of 16ml (~60%) appears to be due to a combination of factors. Once again, difference in the segmentation of disease seems to have been caused due to the deformation in the non-rigidly registered image but also the registration seems to have been affected by changes due to patient motion around the lung. This seems to have caused unwarranted changes on the non-rigidly registered image and has increased TV. In this example, it is the unregistered post-therapy image which has not been able to fully segment all the disease in one attempt and if the post-therapy image has the same areas of disease added the difference in TV is 13.4ml (50%). This is still a sizeable change and likely to be a result of the motion around the lungs causing significant changes in the tumour size and shape on the transformed image. Both of these cases are shown visually in Figure 3.4.

3.4.3.2) Results of SUV and Volume Changes on Pre-Therapy Registered Images

The pre-therapy images in the dataset had double the number of uptake areas compared to the post-therapy scans as there were 58 more tumours before therapy. Table 3.9 shows the mean values for each type of uptake area over the 20 datasets. Once again, rigid registration had very little effect on the SUV_{max} from unregistered pre- therapy images, with no change in all but two of the uptake areas. Two tumours were found to have changes of 0.9 SUV (17%) and 1.8 SUV (11%) between unregistered and rigidly registered scans. SUV_{peak} differences showed consistent small differences in tumours of <0.5 SUV, with one exception of 1.8 SUV. The largest percentage differences were 18.5%, 12% and 11%.

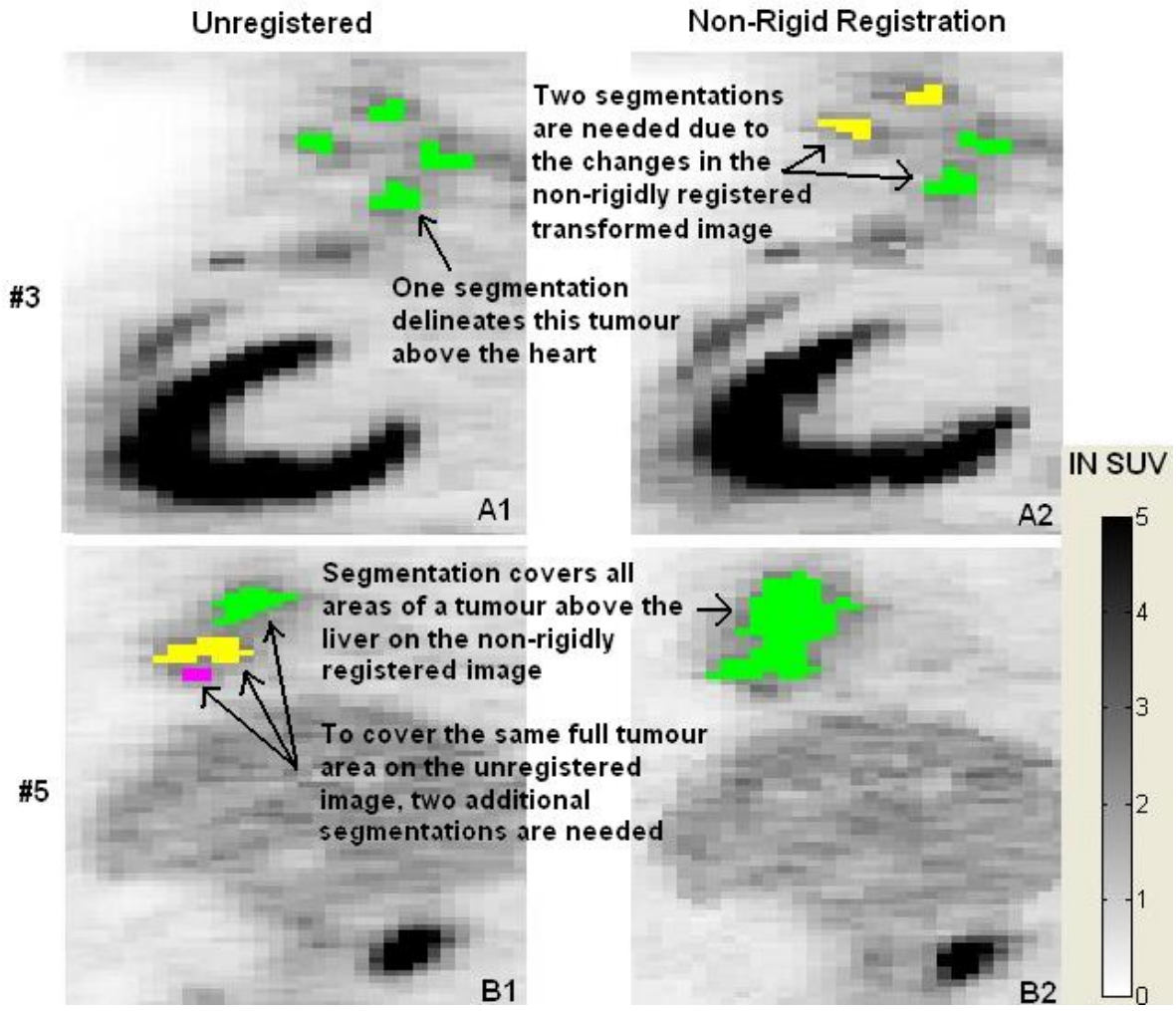


Figure 3.4: Volume Changes between Registered and Unregistered Post-Therapy Images

The images show sagittal slices from (A1) Dataset 3 unregistered post-therapy PET image, (A2) Dataset 3 non-rigidly registered post-therapy PET image, (B1) Dataset 5 unregistered post-therapy PET image, and (B2) Dataset 5 non-rigidly registered post-therapy PET image. For Dataset 3, the non-rigidly registered post-therapy image needed further segmentation to include all the disease seen in the unregistered post-therapy image. For Dataset 5, there appears to be significant deformation on the non-rigidly registered image meaning the tumour has increased in volume and included more disease than in the segmentation on the unregistered post- therapy image. All volumes were segmented using a fixed 2.5 SUV threshold region growing algorithm.

TV for rigid registration showed very little difference between images with an absolute mean difference of 0.23ml (<1%) between volumes. The largest absolute difference in tumours was just 4ml (0.7%) and 95% of tumours had a volume change of <1ml, the equivalent of <12 voxels. Over half the tumour uptake areas showed no difference in volume.

Pre-therapy images with non-rigid registration transformations applied showed significant differences compared to unregistered images for all parameters, particularly SUV_{peak} and volume. There were more substantial changes in SUV_{max} than when using rigid transformations with a mean difference of 2.2% over all uptake areas and 2.5% in tumours. The largest differences in SUV_{max} in tumours were 3.1 SUV (14%) and 1.8 SUV (11%), with other differences <1.5 SUV. The differences of 3.1 and 1.8 SUV were in dataset 2 and dataset 20 respectively and assessment of these tumours revealed that they were positioned in the lungs and images showed more significant change in their visual appearance compared to other tumours, most likely due to deformations in non-rigid registration transformations (Figure 3.5). The tumour highlighted in dataset 20 corresponded to the same tumour which had the largest absolute SUV_{max} change when using the rigid registration transformation. Despite these SUV_{max} changes in some tumours, 73% showed no change in SUV_{max} between the non-rigidly registered pre-therapy PET image and the unregistered equivalent but those that did change were of a significant margin.

		IRTK RIGID REGISTRATION				IRTK NON-RIGID REGISTRATION			
Uptake Area	**No**	**SUV_{max}**				**SUV_{max}**			
		Unreg	**RR**	**Diff**	**% Diff**	**Unreg**	**NR**	**Diff**	**% Diff**
Tumour	62	10.95 (5.67)	10.91 (5.65)	0.04 (0.25)	0.43 (2.43)	10.95 (5.67)	10.74 (5.61)	0.22 (0.53)	2.47 (5.91)
Bladder	19	36.84 23.01)	36.84 (23.01)	0.00 (0.00)	0.00 (0.00)	36.84 (23.01)	35.82 (21.47)	1.02 (2.32)	1.61 (3.33)
Kidney	18	21.35 (31.10)	21.35 (31.10)	0.00 (0.00)	0.00 (0.00)	21.35 (31.10)	21.21 (31.09)	0.14 (0.42)	1.05 (3.85)
Heart	6	10.40 (4.37)	10.40 (4.37)	0.00 (0.00)	0.00 (0.00)	10.40 (4.37)	10.08 (4.52)	0.32 (0.45)	4.69 (7.91)
Total	105	17.39 (19.20)	17.36 (19.21)	0.03 (0.20)	0.25 (1.87)	17.39 (19.20)	17.03 (18.74)	0.35 (1.11)	2.20 (5.34)
Uptake Area	**No**	**SUV_{peak} (27 voxels)**				**SUV_{peak} (27 voxels)**			
		Unreg	**RR**	**Diff**	**% Diff**	**Unreg**	**NR**	**Diff**	**% Diff**
Tumour	62	6.14 (3.26)	6.08 (3.18)	0.08 (0.25)	1.37 (3.41)	6.14 (3.26)	5.56 (2.94)	0.62 (0.81)	10.56 (10.29)
Bladder	19	33.68 (33.52)	33.43 (33.46)	0.39 (0.88)	1.28 (2.68)	33.68 (33.52)	30.70 (33.11)	3.49 (8.27)	9.23 (18.06)
Kidney	18	12.67 (20.11)	12.59 (20.13)	0.07 (0.18)	1.21 (2.64)	12.67 (20.11)	12.54 (20.01)	0.76 (1.89)	4.54 (8.56)
Heart	6	7.31 (3.27)	7.28 (3.20)	0.10 (0.12)	1.57 (1.87)	7.31 (3.27)	6.02 (2.27)	1.29 (1.35)	17.98 (12.07)
Total	105	12.31 (19.36)	12.21 (19.30)	0.13 (0.44)	1.34 (3.06)	12.31 (19.36)	11.33 (18.73)	1.20 (3.75)	9.71 (12.11)
Uptake Area	**No**	**Volume (Fixed 2.5 SUV Threshold Segmentation)**				**Volume (Fixed 2.5 SUV Threshold Segmentation)**			
		Unreg	**RR**	**Diff**	**% Diff**	**Unreg**	**NR**	**Diff**	**% Diff**
Tumour	62	78.94 (189.0)	78.8 (189.7)	0.23 (0.66)	0.73 (1.95)	78.94 (190.0)	63.01 (146.6)	18.26 (46.83)	20.77 (19.05)
Bladder	19	151.2 (139.9)	151.1 (139.9)	0.15 (0.18)	0.11 (0.21)	151.2 (139.9)	190.7 (186.9)	43.76 (63.89)	21.69 (24.70)
Kidney	18	41.22 (27.12)	41.28 (21.12)	0.12 (0.16)	0.29 (0.33)	41.22 (27.12)	40.85 (26.86)	1.56 (1.90)	4.94 (6.82)
Heart	6	216.8 (159.7)	216.5 (159.5)	0.32 (0.36)	0.21 (0.32)	216.8 (159.7)	223.1 (152.7)	18.02 (15.89)	8.61 (6.59)
Total	105	93.42 (167.1)	93.29 (167.1)	0.21 (0.52)	0.52 (1.53)	93.42 (167.3)	91.47 (153.7)	19.20 (46.58)	17.53 (19.23)

Table 3.9: Changes in SUV_{max}, SUV_{peak} and Volume between Unregistered Pre-Therapy Scans and IRTK Registered Pre-Therapy Scans

All parameters are given as the mean for each uptake area over the 20 datasets. SUV_{max} is the intensity of the maximum voxel in the uptake area, SUV_{peak} represents the mean of the SUV_{max} and its 26-connected neighbouring voxels, and the volume (in ml) is segmented using fixed 2.5 SUV threshold region growing algorithm. Unreg = Unregistered pre-therapy values, RR = rigidly registered pre-therapy values, NR = non-rigidly registered pre-therapy values, Abs Diff = absolute difference between unregistered and registered values, Abs Diff % = % difference between unregistered and registered values. (S.D. in brackets).

The mean difference in SUV_{peak} between unregistered and non-rigidly registered post-therapy scans was ~10% across all uptake areas and 10.6% for tumours. For tumours, there was a mean difference of 0.6 SUV and while changes were more significant compared to rigid registration transformations, they were more consistent. Only two differences in SUV_{peak} were >2.5 SUV. SUV_{peak} is more likely to change with registration, particularly non-rigid, as the transformations can mean that the area around the SUV_{max} changes shape and therefore covers different voxels. The two largest differences in SUV_{peak} were in dataset 20, although in different tumours to the largest SUV_{max} changes. These changes suggest that registration of dataset 20 resulted in a transformation which significantly altered the registered PET image. This could be due to greater reductions in tumour size between pre- and post- therapy CTs and/or the positioning of tumours in and around the lung where breathing is likely to cause more changes between pre- and post-therapy images. This can result in more deformation in non-rigid transformations in an attempt to align images causing values to be affected more.

Volumes of the uptake areas showed substantial differences between unregistered and non-rigidly registered pre-therapy images, particularly in tumours where differences of >100ml were found in 10% of tumours. There was a mean difference of 18ml (21%) between TVs. In contrast to differences in post-therapy images, where areas of disease had been under- or over- segmented (Figure 3.4), differences in pre- therapy images appear to be because of a clear change in TV (Figure 3.6). This is most likely because of greater deformations of shape and manipulation of values on the transformed PET image caused by the non-rigid registration algorithm. There were six TVs which changed by >100ml and all of these were due to the volume shrinking causing significant changes in tumour size.

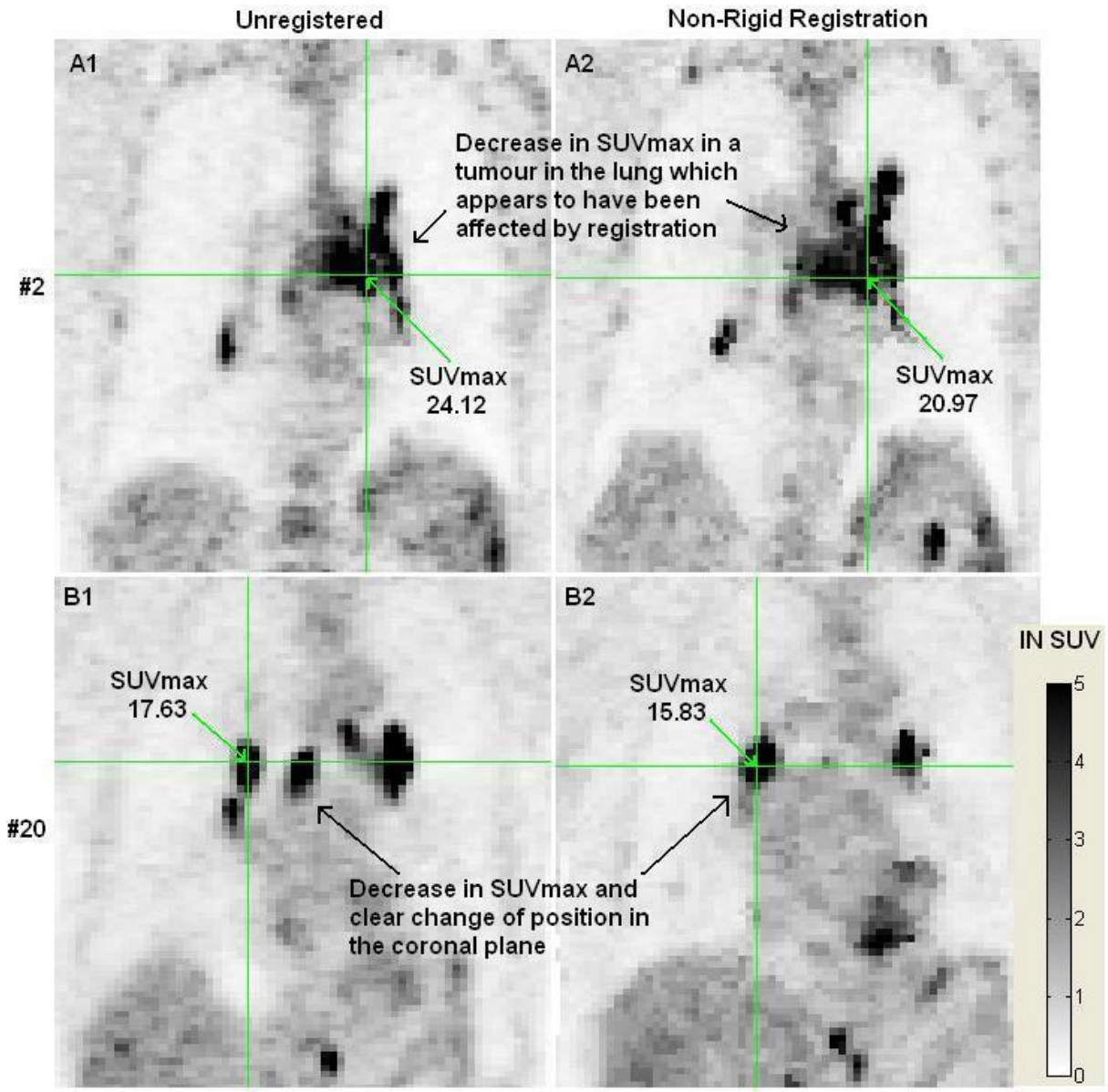


Figure 3.5: Changes in SUV_{max} between Unregistered and Registered Pre-Therapy Images

The images show coronal slices from (A1) Dataset 2 unregistered pre-therapy PET image, (A2) Dataset 2 non- rigidly registered pre-therapy PET image, (B1) Dataset 20 unregistered pre-therapy PET image, and (B2) Dataset 20 non-rigidly registered pre-therapy PET image. Both show the different positions, highlighted by green crosshairs, and values of SUV_{max} between images. As can be seen, particularly in Dataset 20, different coronal slices contain the SUV_{max}.

Figure 3.6 highlights two of these differences in TV on dataset 6 (reduction of 475ml to 278ml) and dataset 15 (reduction from 607ml to 496ml). The non-rigid registration on dataset 15 also seems to have caused a registration artefact visible in the lower half of the spleen. Given these examples and the overall results (Table 3.9) it is clear that non-rigid registration causes significant changes in TV and a modest reduction in SUV_{max} and SUV_{peak}. There was a mean TV of 79ml on unregistered pre-therapy images compared to a mean of 63ml on non-rigidly registered pre-therapy images. All changes in SUV_{max} showed a reduction from the unregistered pre-therapy images to those with non-rigid registration applied and while there were some increases in SUV_{peak} and TV, the significant majority of changes were negative.

Registration transformations on both pre-therapy and post-therapy images show a mean decrease in SUV_{max} and SUV_{peak} values, potentially due to interpolation effects on a transformed image. While transformed post-therapy images show a mean increase in TV (20.8ml to 23.1ml), transformed pre-therapy images show a mean decrease in TV (79ml to 63ml). This is to be expected as there is greater tumour mass on pre-therapy CT images compared to post-therapy CT images, where treatment has reduced tumour size. Non-rigid registration will deform post-therapy CT TVs to try and match larger volumes on pre-therapy scans. In the reverse, registration of pre-therapy CT scans will deform tumour size and shape to match smaller lesions on post-therapy scans. Therefore, when these transformations are applied to pre-therapy PET images it is unsurprising that there is a reduction in the size of segmented TVs from unregistered PET images. Rigidly registered images, which only allow the translation and rotation of tumours, do not suffer from the same issues with changes in TV as they do not manipulate the size or shape of the tumour, so only minimal changes between transformed and unregistered images are found.

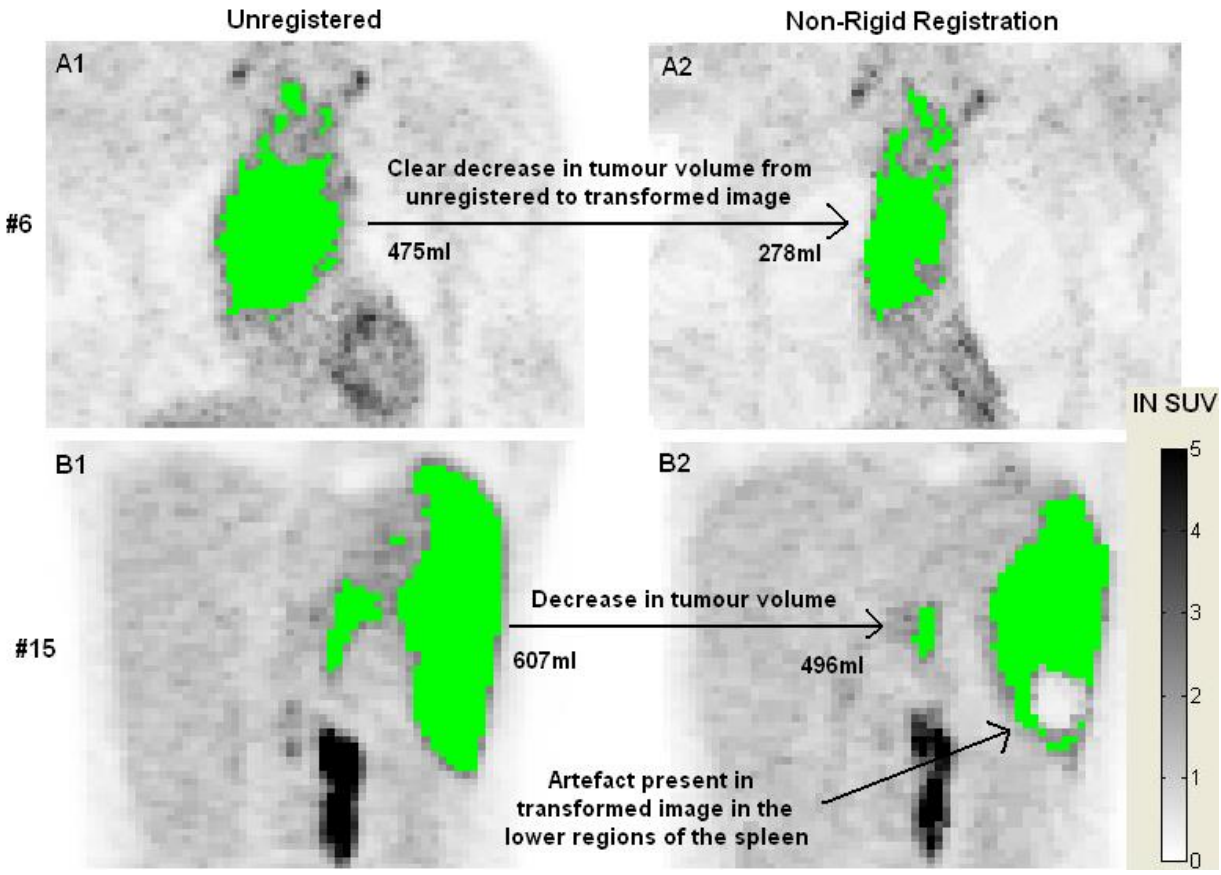


Figure 3.6: Volume Changes between Unregistered and Registered Pre-Therapy Images

The images show coronal slices of (A1) Dataset 6 unregistered pre-therapy PET image, (A2) Dataset 6 non-rigidly registered pre-therapy PET image, (B1) Dataset 15 unregistered pre-therapy PET image, and (B2) Dataset 15 non-rigidly registered pre-therapy PET image. For both datasets, there is a significant decrease in the segmented TV due to the non-rigid registration transformation reflecting the volume changes in the corresponding CT images. In dataset 15, the registration has also caused an artefact, visible in the lower spleen. All volumes were segmented using a fixed 2.5 SUV threshold region growing algorithm.

There was no noticeable relation between changes in SUV and volume of PET tumours in the pre-therapy images and the size of pre-therapy CT lesions or change in size of CT lesions from pre- to post- therapy. For each dataset, mean pre-therapy CT tumour size and relative change in RECIST tumour measurements were correlated with the mean percentage differences in SUV_{max}, SUV_{peak} and TV for both rigid and non-rigid registration. Relative change was given as the

product of absolute change and percentage change between the pre- and post- therapy CT measurements. Pearson correlation coefficients (pcc) were calculated and showed just one significant correlation of the 24 tested (Table 3.9). This was done per dataset rather than for each individual tumour measurement as there was considerable difficulty in establishing which CT measurements related to which PET tumours. Anatomical and metabolic areas of disease can be distinct as metabolic activity and change in intensity do not necessarily relate to the change in anatomical size of a significant area of tumour mass. Equally, an area of tumour mass may warrant a different number of measurements on each imaging modality – a PET segmentation may include two or three lesions on CT or vice versa. Correlation was performed on 17 of the 20 datasets. Two were omitted because there were no viable PET tumour measurements that could be made and another was omitted due to no viable CT measurements.

In relation to the number of datasets, correlation values would have to be at least 0.483 to be significant ($p < 0.05$). There is a weak but insignificant correlation between relative changes of CT lesions and differences in SUV_{max} and SUV_{peak} when using non-rigid registration (pcc = 0.365, $p < 0.15$, and pcc = 0.325, $p < 0.21$, respectively). This suggests that greater change in CT tumour size increases the likelihood of there being changes in SUV between unregistered and non-rigidly registered scans. There is a correlation between the pre-therapy CT size and the difference in TV between unregistered and rigidly registered images (pcc = 0.512, $p < 0.036$). However, this is not unexpected as a larger pre-therapy CT size is likely to result in greater absolute differences in segmentation as there is more disease to segment and therefore greater chances of differences. Non-rigid registration is likely to cause more unpredictable changes, compared to rigid registration and therefore does not have similar correlation.

Pearson Correlation Coefficients				
Change in	**CT Size**		**CT Relative Change**	
	RR	**NR**	**RR**	**NR**
SUV_{max}	-0.163	-0.04	0.040	0.365
SUV_{max} (%)	-0.120	-0.06	0.119	0.238
SUV_{peak}	-0.270	-0.01	-0.113	0.325
SUV_{peak} (%)	-0.259	0.03	0.008	0.284
Tumour Volume	0.512	0.10	-0.100	-0.104
Tumour Volume (%)	-0.344	0.22	-0.163	0.056

Table 3.9: Pearson Correlation Coefficients between CT Size and SUV/Volume

Pccs for the difference in SUV_{max}, SUV_{peak} and TV on unregistered and registered pre-therapy images, using both IRTK rigid (RR) and IRTK non-rigid (NR) algorithms, correlated with the pre-therapy CT tumour size as measured by RECIST criteria (Table 3.7) and relative change between the pre- therapy and post- therapy CT tumour size. Correlations were calculated for absolute and percentage differences of SUV_{max}, SUV_{peak} and TV.

The exact reason for changes in transformed registered images has not been investigated in the literature. In performing any sort of registration there are likely to be interpolation effects from transformation parameters covering changes on a registered image. For a rigid algorithm, where the registration only changes the translation and rotation of an image, this is unlikely to have much of an impact, apart from at the edge of images, especially in comparison to non-rigid registration. A non-rigid algorithm will produce a more complicated transformation and is more likely to affect the intensity values in an image and the shape of high intensity areas, depending on the properties of the chosen transformation. Therefore, the TVs segmented in these PET images are likely to change. Changes in segmented PET TV occur not just because of changes in intensity but also due to changes in the size and shape of tumours. The results obtained when looking at the differences between unregistered and registered images do correlate with these theories.

3.4.4) Summary of Results for SUV and Volume Changes

The application of IRTK rigid and non-rigid registration transformations to PET scans can cause differences from unregistered images. Rigid registration causes minimal changes to SUV_{max} while SUV_{peak} and TV show very small differences with a mean change of <2%. Non-rigid registration transformations can cause large changes in tumours from unregistered images with SUV_{max} differing by a mean of 2%, SUV_{peak} by 11% and TVs by 21%. These results suggest that rigid registration transformations can be applied with little concern over changes between unregistered and registered images, however, non-rigid registration transformations must be used with care as they can cause large changes in SUV and TV which may cause issues when analysing registered pre- and post- therapy images.

3.4.5) Theory and Importance of Results for SUV and Volume Changes

One of the reasons for registering pre- and post- therapy scans is to obtain subtraction images which can be used to identify response to therapy. Parametric images have been used in other studies and extracted parameters have been shown to predict response (Necib *et al*., 2008; Necib *et al*., 2011). In an ideal scenario, with perfect image registration and unaffected transformed images, this methodology is, theoretically, an excellent way of investigating changes in tumours and predicting response. However, in reality, there are many issues which cause this technique to be problematic.

The methodology used for registering pre- and post- therapy PET/CT studies uses the CT component of the PET/CT scan to register the images. The reason for this is that the CT has a higher resolution and there is less change in morphology between scans than on PET (Weider *et al*., 2005). However, CT registration relies on minimal change in anatomy between pre- and post-therapy scans. Otherwise, changes in tumour size can affect the registration accuracy and

resulting transformation. Data has shown that the reduction in CT tumour size from pre- to post-therapy scans is of a significant degree and is therefore likely to affect the registration (Table 3.7). Additionally, registration on an image is unlikely to be perfect on a voxel-by-voxel basis, no matter how impressive the algorithm used. Even when applying the CT transformations on a PET image, which has a lower resolution, matching two PET scans to a voxel level accuracy is difficult to achieve and clearly vital for parametric imaging to be successful.

Using transformed, registered images also poses another issue as the application of the transformation on an image can result in changes in SUV and of the shape and size of tumours (Table 3.8; Table 3.9). A subtracted image should be of the pre-therapy SUVs subtracted by the post-therapy SUVs, however, if the post-therapy SUVs have changed due to the registration process then this has an effect on the subtraction image, as this is not a true representation of the difference between the two images. These three issues mean that a parametric image is unlikely to detect just the changes in tumours between pre- and post- therapy scans but also the error in the registration algorithm, the affect of changes in tumour size on CT and the affect of the transformation on the post-therapy PET scan.

Figure 3.7 attempts to depict the difference between the idealistic scenario, in which parametric imaging is almost certainly of great use, and the realistic scenario, in which the aforementioned issues make parametric imaging problematic. Figure 3.7 (A1, A2) depict a theoretical pre-therapy PET/CT image with a homogeneous tumour. In an ideal scenario, the tumour in the post-therapy PET image responds to treatment and shows a metabolic change, i.e. a reduction in size and intensity, while the size of the tumour on CT remains the same (Figure 3.7 (B1, B2)). The stable size of the tumour on the CT image aids the registration of pre- and post- therapy PET/CT images, so registration is perfect, and the transformed post-therapy PET image suffers no changes in tumour SUV or volume from the unregistered PET image (Figure 3.7 (C1, C2)). However, in a

realistic scenario, the size of the tumour on the CT post-therapy image is likely to reduce in size, as well as change position (Figure 3.7 (D1, D2)). This is likely to cause more issues in the registration, which is not going to be perfect, and SUV and tumour size are likely to change between the unregistered and transformed post- therapy image too (Figure 3.7 (E1, E2)).

The difference between the ideal and realistic scenarios does not mean that there is no hope for parametric imaging. However, these flaws do need to be considered and addressed as best possible to make it a success. To see if this technique is a plausible method of identifying response to therapy, it makes sense to use data with minimal change in the size of CT lesions between pre- and post- therapy scans. Equally, the registration method used needs, if possible, to be one which is both accurate and preserving of SUVs when the transformation is applied to a post- therapy image.

Unfortunately, the data used herein has large changes in the size of tumours on CT from pre- to post- therapy images (there is a mean reduction of 16mm (58%) in the longest diameter). It may be that this is typical across the majority of datasets. However, it could be that some studies, with different scanning times and types of disease, have much smaller changes. While the registration algorithms used in this work were shown to be accurate by both qualitative and quantitative analysis, there is the possibility that other methods could enhance the robustness and accuracy of registration further. The affect on SUV may be dependent on the data and registration methods used and the affect of them on the change in SUVs between images should be considered.

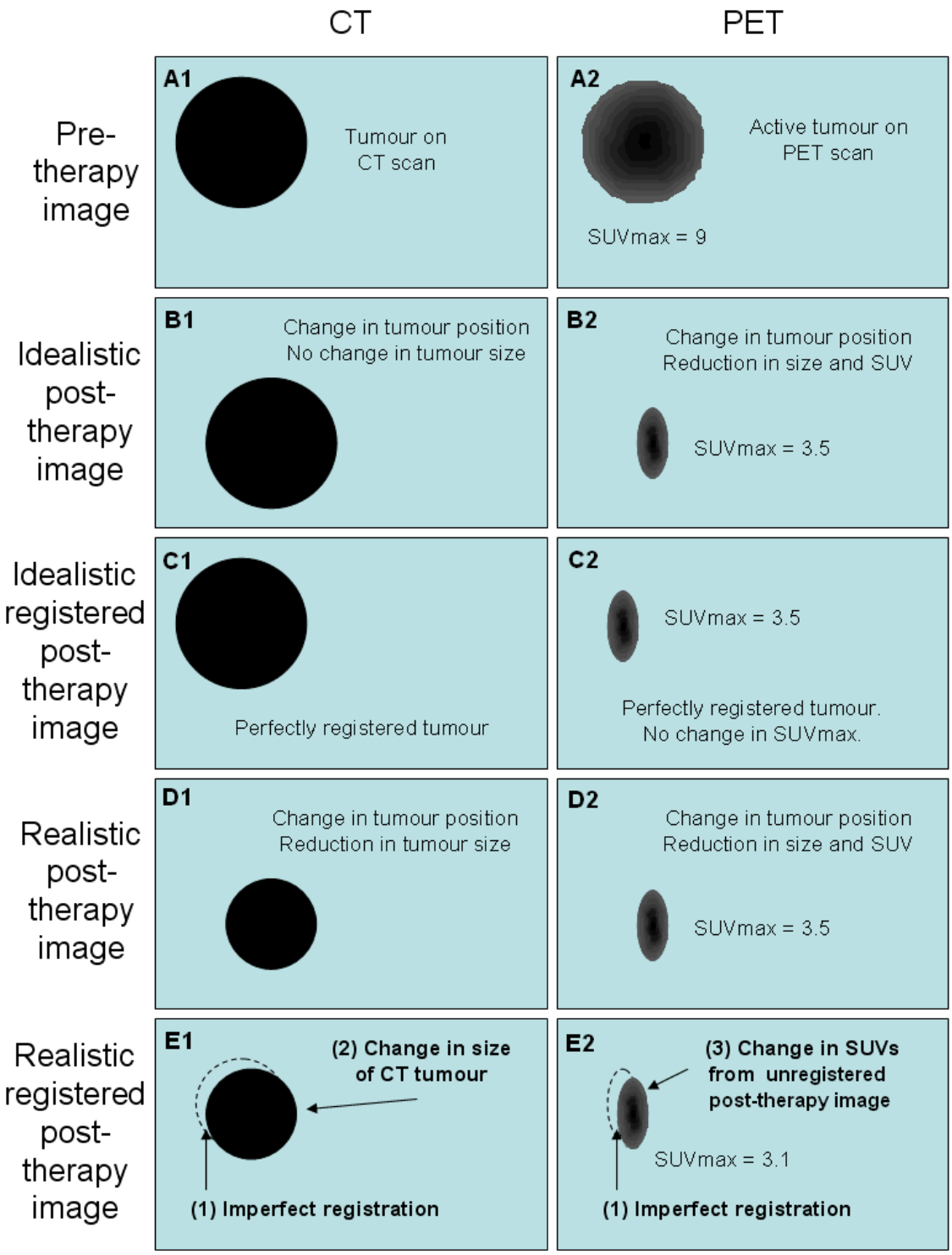


Figure 3.7: Theoretical Issues with using Registered Post-Therapy Images in PET Analysis

Illustration of the potential issues in registering pre- and post- therapy PET/CT scans for parametric imaging in an attempt to identify response to therapy.

To increase the chances of parametric imaging working, a dataset which has minimal changes in tumour size between scans on CT and a registration method which is both accurate and reduces the change in SUV on transformed images is needed. In terms of registration accuracy and the change in SUV on transformed images, there becomes a possible trade-off in what is more important. Results show that the non-rigid algorithm is more accurate than the rigid algorithm making it more suitable for parametric imaging. However, the non-rigid algorithm also shows larger changes in SUV on transformed post- therapy images so which of these registration methods is more suitable for parametric imaging is dependent on which of these factors is more important and how significant the effect is on the analysis. There is also more to understand in terms of how the registration transformation parameters manipulate the image. It is possible that other affects could also cause issues in the process of using parametric images, for example, the presence of artefacts on the transformed image (Figure 3.6) should theoretically ruin any attempt at predicting response using parametric images.

3.5) Discussion on Registration of PET/CT Scans

Herein, a registration methodology was used that registers pre- and post- therapy PET/CT scans, using the CT component of the scan. The methodology used both customised rigid and non-rigid IRTK algorithms to align CT images, with the resulting transformations applied to PET images. The rigid and non-rigid registrations were found to have a registration accuracy of ~10mm and ~6.5mm respectively on a quantitative landmark analysis on CT and had good visual scores in a qualitative analysis on both PET and CT images. These results suggest both types of registration are useful in aligning PET/CT studies whether it is for visual assessment or quantitative purposes.

However, there are issues with using both types of registration for aligning pre- and post- therapy PET/CT images. The rigid registration does not produce the accuracy of alignment in comparison

to non-rigid registration meaning voxel-by-voxel analysis is flawed. Without tumour areas overlapping accurately, this type of analysis will become redundant as results will predominantly show differences in mis-registration rather than changes in the tumour. There is the possibility that rigid registration applied to smaller volumes containing tumours will produce more accurate registration which could be suitable for voxel-by-voxel analysis techniques but this would need thorough investigation.

Non-rigid registration suffers from the opposite problem to rigid registration. It is likely to register images very well but the danger is that it does so by changing the source image in the registration too much in order to align changes in tumour size on the CT, particularly if large changes on the CT are seen between pre- and post- therapy scans. There are significant changes in tumour size between pre- and post- therapy CT scans in the dataset used for registration validation (Table 3.7). This could result in a non-rigid algorithm deforming a post-therapy image to match a pre-therapy image with changes being unrealistic and deforming the image too much to be considered realistic registration. This is then transferred onto the PET image. It would be useful to discover if non-rigid registration would have a similar affect on PET images if tumours on CT images show little change between scans, which could be the case in other types of datasets with different diseases. A major factor in the registration process with this methodology is the assumption that the change between CT lesions from pre- to post- therapy images will be smaller than the changes on PET images. While this is the case, RECIST measurements show that CT changes were of a larger factor than expected and this may have unwanted effects on transformed images, in particular, when using the non-rigid transformations where deformations will have been applied to match changes in tumour size. This is supported by results analysing the affect of registration transformations which produced changes of 21% on TV and 11% on SUV_{peak} between unregistered and non-rigidly registered scans.

Using registrations to analyse post-therapy images becomes a trade-off between using a rigid registration, which is unlikely to produce large changes from unregistered image values but will have more imperfect alignment in comparison to non-rigid registration, and non-rigid registration, which will produce more accurate image alignment but is likely to manipulate image values in a way which could be damaging to any analysis. Both methods are worth further investigation and show accuracy which suggests they may be useful in analysing response to therapy in pre- and post- therapy PET/CT scans. However, the inherent flaws of both must be considered during analysis. Other registration methods which could provide more accurate registration without deforming original image values would be of particular interest.

3.6) Conclusion to Registration of PET/CT Scans

Registration of PET/CT scans is a relatively new area of research and data presented herein has shown there is potential for accurately registering scans. In a cohort of 20 lymphoma patients misalignment errors between pre- and post- therapy images were found to be 40mm on average, which is greater than the size of most lesions of interest in patients. Image registration was shown to substantially reduce this misalignment with both rigid and non-rigid IRTK registrations found to be accurate methods for registering pre- and post- therapy PET/CT images. Both rigid and non-rigid IRTK registrations obtained good visual analysis results on both PET and CT scans and demonstrated an average misalignment error of ~10mm on rigid and ~6.5mm on non-rigid images using an anatomical landmark study on CT data.

The affect of the registration transformations on the PET images was also investigated. Rigid registration caused minimal changes in SUV_{max} values while differences in SUV_{peak} and TV were minimal with mean a percentage difference of <2%. Non-rigid registration transformations caused more significant changes, with SUV_{max} differing by a mean of 2% and SUV_{peak} and TV varying by differences of 11% and 21% respectively between unregistered and non-rigidly

registered images for tumours. While non-rigid registration was found to be more accurate, the substantial changes it showed in the transformed image's SUV and TV suggest that using transformed images to identify response is problematic. On the other hand, while rigid registration produced transformed PET images with little or no change, an estimated registration accuracy of ~1cm may be too large for voxel-by-voxel analysis to be applicable.

4) Response to Therapy in Patients with Mesothelioma

4.1) Introduction to Response to Therapy in Patients with Mesothelioma

4.1.1) Identifying Response to Therapy in Mesothelioma Patients

The aim of this study was to use PETTRA and the response parameters it can produce to analyse a group of patients with mesothelioma. All patients in the study underwent pre- and post- therapy PET/CT scans. Images were analysed using PETTRA software with the aid of an experienced clinician to deduce which areas of uptake were disease, and which were physiological. Image analysis response parameters were compared with measures of survival to deduce their effectiveness. Comparisons were made between different response parameters and segmentation methods to deduce their variability and reproducibility.

4.1.2) Mesothelioma

Malignant pleural mesothelioma (MPM) is a rare cancer of the pleura that arises from cells which then develop into either epithelial or sarcomatoid neoplasms (Rusch, 1995). The involvement of epithelial, sarcomatoid or mixed cells determines the subtype of disease with epithelial subtypes tending to survive longer (Bénard *et al*., 1999). Tumours spread locally, typically extending into thoracic structures such as chest wall, pericardium and myocardium (Bénard *et al*., 1998). Patients commonly die because of cardiac or pulmonary involvement after persistent growth of disease, at which point blood-borne metastases can occur and are often detected at autopsy (Roberts, 1970; Sugarbaker *et al*., 1996). Patients are usually diagnosed after experiencing symptoms such as dyspnoea (shortness of breath), chest pain, cough or weight loss (Wang *et al*., 2004). MPM is often diagnosed late as the disease develops within body cavities and patients can often have extensive tumour involvement by the time they are diagnosed (Robinson and Lake, 2005).

Mesothelioma has been strongly linked to asbestos exposure, up to 30-40 years previously, and as a result mesothelioma incidence is expected to rise as the worst affected cohorts were born in the 1940s (Peto *et al*., 1999). A recent study estimates mesothelioma causing up to 250,000 deaths in Western Europe from 1995-2019, peaking in incidence between 2015 and 2019 (Pelucchi *et al*., 2004). In Great Britain, mesothelioma deaths are predicted to peak between 2011 and 2015, causing between 1950 and 2450 deaths (Hodgson *et al*., 2005). Many cases in South East England occur along the Thames and its estuary, where shipbuilding and other industries used asbestos within the last century (Mak *et al*., 2008). Asbestos is thought to cause mutations in some of the estimated two billion mesothelial cells in the mesothelium (Robinson and Lake, 2005), while the simian virus (SV40) is also believed to be a factor, particularly in the rarer cases where there is no asbestos exposure, as mesothelial cells have high levels of p53 protein expression and are unusually susceptible to SV40 mediated transformation (Gazdar and Carbone *et al*., 2005). However, asbestos remains the major cause and the role of SV40 remains unclear (Robinson *et al*., 2005). MPM is most common in older men, ~80% of patients are male and the age of patients at diagnosis is typically ~60 (Herndon *et al*., 1998; Andreopoulou *et* al., 2004; Bottomley *et al*., 2006). The disease is usually fatal with a median survival of <12 months even with chemotherapy (Andreopoulou *et al*., 2004; Bottomley *et al*., 2006; Schaefer *et al*., 2012), although, on occasions, median overall survival can be doubled with effective treatment (Gerbaudo *et al*., 2011).

4.1.3) Treatment of Mesothelioma

Treatments vary for patients with MPM, ranging from radiotherapy for palliation only or combined modality treatment involving chemotherapy, aggressive surgical resection and radiotherapy. Surgery can be both palliative, for example, partial pleurectomy with pleurodesis to control effusions, or curative, involving an extrapleural pneumonectomy or radical pleurectomy and decortication to try and remove all gross tumour in the patient (Sugarbaker *et al*., 2004). Only

a few patients are suitable for curative surgery and adjuvant therapy is usually considered afterwards to eliminate any residual disease (Robinson *et al*., 2005). Early reports suggested unacceptable levels of toxicity involved with the use of radiotherapy in MPM, however, radiotherapy is often used locally, following curative surgery and for palliation of symptoms (Baldini, 2009). The use of advanced radiotherapy methods, including intensity-modulated radiotherapy (IMRT), may reduce toxicity by more accurately irradiating the desired areas in patients, thus making it a more attractive treatment option (Baldini, 2009).

Several chemotherapy regimes have proved valuable for palliation in MPM patients, reducing tumour burden, pain and breathlessness. Pemetrexed plus cisplatin, with a response rate of 41% and an increase in survival of three months in comparison to chemotherapy with just cisplatin (Vogelzang *et al*., 2003), and gemcitabine plus cisplatin, with a response rate of 48% (Nowak *et al*., 2002), have proven to be the most effective. While surgery, radiotherapy and chemotherapy have yielded disappointing survival times on their own, studies have shown that selected patients who have surgery and then undergo chemotherapy and/or radiotherapy have an improved median survival time of ~18-19 months (Sugarbaker *et al*., 1999; Lee *et al*., 2002). Pemetrexed plus cisplatin chemotherapy is commonly used as standard first line treatment in many institutions with no standard second line treatment yet established (Hazarika *et al*., 2005).

Many chemotherapy regimes have been used as second line treatment including vinorelbine, pemetrexed plus carboplatin, raltitrexed plus oxaliplatin, oxaliplatin plus carboplatin, gemcitabine plus vinorelbine, ZD0437 (a platinum analogue) and a combination of irinotecan, cisplatin and mitomycin-C with disappointing response rates of <20% (Ceresoli *et al*., 2010). MPM has been shown to express vascular endothelial growth factor (VEGF) which could be a potential drug target (Ohta *et al*., 1999). Using targeted therapy in pre-treated MPM patients has become more popular, with studies investigating the impact of agents such as thalidomide, sorafenib, sunitinib,

belinostat, erlotinib plus bevacizumab, and ranpirnase plus doxorubicin (Ceresoli *et al*., 2010). Sorafenib is an inhibitor of Raf-kinase, VEGF receptor-2, and platelet-derived growth factor receptor-b, another target of MPM, making it a candidate for second line treatment (Wilhelm *et al*., 2004). A phase II trial using sorafenib to treat 51 mesothelioma patients reported a poor response rate but good clinical benefit and increased survival times (Janne *et al*., 2006).

4.1.4) Role of ^{18}F-FDG PET/CT in Management of Mesothelioma

The pattern of growth of MPM makes it a challenging disease to quantify using traditional, anatomical CT imaging. Standard RECIST criterion has been found to be inadequate for measuring MPM tumours (van Klaveren *et al*., 2004). Updated RECIST criteria have been created for MPM patients (Byrne and Nowak, 2004), and used to predict survival in MPM patients treated with chemotherapy (Schaefer *et al*., 2012). However, it is the PET component of PET/CT which is of the most interest in recent mesothelioma studies as its functional imaging capabilities provide metabolic information anatomical imaging modalities cannot obtain. PET/CT is used in patients with MPM for pre-operative staging, post-treatment surveillance and response to treatment assessment (Basu *et al*,. 2011). In patients with MPM, high SUVs have been shown to correlate with poor survival (Flores, 2005; Ceresoli *et al*., 2006). SUV is also higher in malignant lesions than in benign pleural lesions related to asbestos exposure (Bénard *et al*., 1998; Lee *et al*., 2009; Sharif *et al*., 2011). However, variation of SUVs in patients with MPM can be large (Bénard *et al*., 1999), and represent only a single pixel rather than the true extent of the disease. PET tumour volume (TV) and total lesion glycolysis (TLG), which take into account the entirety of the tumour and the degree of uptake within that entirety, could be more relevant measures.

4.1.5) Identifying Response to Therapy in Mesothelioma Patients

TV and TLG are theoretically ideal measurements for assessing response in MPM. MPM can be an expansive and heterogeneous disease (Figure 4.1), unlike other cancers with smaller, more homogenous tumours. Therefore, analysing the full extent of the disease could be more beneficial than measuring its maximum intensity. SUV_{max} or SUV_{peak} measurements do not show the change in spread of disease and will not accurately reflect changes in tumour distribution between pre- and post- therapy scans, whereas TV and TLG will account for this. TLG has been found to relate to clinical results and survival better than SUV_{max} in a number of studies (Steinhart *et al*., 2005; Francis *et al.*, 2007; Lee *et al*., 2010; Nowak *et al*., 2010; Veit-Haibach *et al*., 2010; Schaefer *et al*., 2012). TV has been segmented using different algorithms including 2.5 SUV thresholding to account for all voxels within a cubic VOI placed over the disease (Veit-Haibach *et al*., 2010; Schaefer *et al*., 2012), the recommended segmentation in the PERCIST criteria (Lee *et al*., 2010), and the GRAB segmentation algorithm (Francis *et al.*, 2007; Nowak *et al*., 2010).

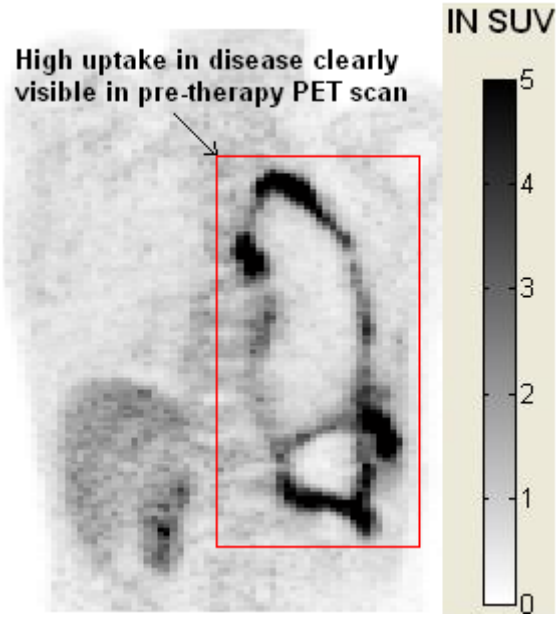


Figure 4.1: PET Scan of a Patient with Mesothelioma

PET scan shows ^{18}F-FDG uptake in mesothelium in the left lung of the patient. The heterogeneity and volume of uptake suggests measures such as TV and TLG could be suitable to quantify the disease.

4.1.6) Aim of Response Analysis

The aim of this chapter was to explore a variety of methods in a pilot study to assess the response to therapy in a cohort of mesothelioma patients treated in a clinical trial in a uniform way with VEGF inhibitor, sorafenib. The primary aim was to investigate whether visual assessment, SUV_{max} or volumetric measures, such as TV and TLG, could predict response. The secondary aim was to investigate different methods of segmentation for obtaining TVs and further methods of quantitative analysis with a view to assess response. Quantitative measures such as TV, TLG and IVH parameters have all shown potential for assessing response (Francis *et al*., 2007; El Naqa *et al*., 2009a, EL Naqa *et al*., 2009b, Lee *et al*., 2010).

4.2) Patients and Scanning

4.2.1) Patient Eligibility and Treatment

53 eligible patients were recruited for the study, 14 of whom had pre- and post- therapy PET/CT scans as part of a subset of patients recruited to a PET pilot substudy, an exploratory study to assess PET/CT as a measure of response. All patients had measurable disease, as defined by modified RECIST criteria (Byrne and Nowak, 2004), a European Cooperative Oncology Group (ECOG) performance status score of 0-2 (Oken *et al*., 1982), a life expectancy of >12 weeks and adequate organ function. All patients were deemed unsuitable for surgery and had received first line treatment of pemetrexed plus cisplatin chemotherapy, before undergoing second line sorafenib chemotherapy. Prior surgery, before relapse, and palliative radiotherapy were permitted. The study was approved by the UK National Research Ethics Service. All patients signed consent forms before the start of treatment which involved continuous dosing with 800mg sorafenib daily in two doses. Dosing was reduced to 400mg daily or every two days if toxicity

was deemed to be an issue. Treatment was continued until disease progression could be confirmed, the patient withdrew from the study or toxicity was deemed to be unacceptable.

4.2.2) PET/CT Scanning

All patients in the study underwent stand alone CT scans at baseline and at 8-weekly intervals as part of the study. Pre- and post- therapy PET/CT scans were acquired at baseline and after therapy for 14 patients between October 2008 and December 2009 for the PET pilot substudy. All scans were acquired at the PET Imaging Centre in St Thomas' Hospital on either a GE Discovery ST or GE Discovery VCT PET scanner (Waukesha, WI). The post- therapy PET/CT scans was performed ~8 weeks after the start of treatment to investigating the success of sorafenib chemotherapy (Papa *et al.*, 2013). For the PET and CT scan components of the PET/CT study, patients were scanned during free breathing and CT scans were low dose non-contrast enhanced scans.

For all 14 PET/CT datasets, CT image dimensions were 512 x 512 x 223 or 267 (with voxel sizes of 0.98mm x 0.98mm x 3.27mm) while PET image dimensions were 128 x 128 x 223, 267 or 311 (with voxel sizes of 5.47mm x 5.47mm x 3.27mm for all but one image with voxel sizes of 4.69mm x 4.69mm x 3.27mm). All PET images analysed were attenuation corrected using a smoothed CT dataset. Administered FDG dose ranged from 315 to 380MBq (median, 342MBq). The dose for the post- therapy scan was within 10% of the dose of the pre- therapy scan for all but one dataset, where the dose was 15% (48MBq) less. The median time between administration of FDG and the start of the scan was 93min (range, 79-122min), with 9/14 pre- therapy scans within 10min of the post- therapy scans. The other five datasets had differences between 12-23min (median: 20min).

4.2.3) Patient Characteristics

Patients had a median age of 63 (range, 55-77) and 86% were male (12 male, 2 female). Patients weighed between 47-99kg (median, 78kg) and there was a mean decrease in weight of 5kg between pre- and post- therapy scans (range, 0-10kg), most likely due to the effects of the disease and the chemotherapy. Patient height ranged from 163-188cm (median, 175cm) with a median change in height between scans of 1cm (range, 0-5cm) within the error of measurement. Of the 14 patients, 13 had stopped treatment at the time of data collection with treatment lasting between 42-300 days (median, 155 days).

4.2.4) Study End Point

When trying to assess a method of predicting response to therapy, an end point or measure was needed to see how successful it is at achieving this goal. A number of different end points can be used as measures of response (Table 4.1), including overall survival (OS) and progression free survival (PFS). OS measures the time till death after treatment has begun, making it a good measure for trials where there is a low response in patients, as would be expected in patients with MPM. PFS is similar to OS except that it measures the time till progression of disease is confirmed, indicating that treatment has failed. In literature, both these methods have been used to measure response in patients with mesothelioma (Benard *et al*., 1999; Hazarika *et al*., 2005; Bottomley *et al*., 2006; Lee *et al*., 2009; Nowak *et al*., 2010; Veit-Haibach *et al*., 2010).

End Point	Definition
Overall Survival (OS)	The time from the entry onto the clinical trial until death as a result of any cause
Progression Free Survival (PFS)	The time from the entry onto a study until progression or death as a result of any cause
Event-Free Survival (EFS)	The time from the entry onto the study to any treatment failure including disease progression or discontinuation of treatment for any reason
Time to Progression (TTP)	The time from the entry onto the study to documented progression or result of death due the disease under study
Disease-Free Survival	The time from the occurrence of a disease-free state or a complete response to disease recurrence or death as a result of the disease under study or toxicity to treatment

Table 4.1: Definitions of End Points for Response Studies

Comparison of different end points used in studies for identifying response in patients.

(Modified from Cheson *et al*., 2007)

4.3) Data Analysis – Response Measures

4.3.1) Visual Analysis

Visual analysis of pre- and post- therapy PET/CT studies were completed by two consultants, a nuclear physician and a radiologist with 20 and 7 years of experience of PET, respectively. Response was assessed using both PET and CT, viewed on HERMES Hybrid Viewer workstations (Nuclear Diagnostics AB, Stockholm, Sweden). Visual assessments of response were scored based on newly developed criteria, defined a *priori* for the trial, categorising patient response as either: response, PMD or SMD (Table 4.2). SUV_{max} values were also listed by consultants as a guide. Scans were viewed scaled to a SUV_{max} of 10, normalised for injected activity and body weight. PET/CT scans were read independently by the two observers and any differences were resolved by consensus. Anatomical visual analysis on CT scans was also performed using modified RECIST for MPM by an experienced radiologist without reference to

the PET/CT scans or reports (Byrne and Nowak, 2004). Patient response was categorised as CR, PR, SD or PD (Table 4.3). A key change in the modified RECIST criteria for MPM is the way tumours are quantified, as measuring the 'longest unidimensional diameter' of disease can be difficult as the tumour mass will often encapsulate the curve of the chest wall. The modified criteria recommend tumour mass is measured in two positions at three separate points at least 1cm apart on transverse slices of the CT, related to anatomical points to increase reproducibility. The sum of these measurements define a pleural unidimensional measure and all the measures combine to make the total tumour measurement.

Response	Cat	Description
Response	A	Overall reduction in intensity of uptake compared to staging scan
	B	Regression of lesions (with or without overall reduction)
PMD	A	Increase in intensity of uptake compared to staging scan
	B	New lesions within affected hemithorax (with or without increased intensity)
	C	New lesions outside affected hemithorax (with or without increased intensity)
SMD		Response does not fall into response or PMD categories

Table 4.2: Criteria for PET Visual Assessment of Patients with Mesothelioma

Response to therapy can be defined as response, progressive metabolic disease (PMD) or stable metabolic disease (SMD) with separate categories (Cat) defining response and PMD.

Response	Description
CR	Disappearance of all target lesions with no evidence of tumours elsewhere
PR	>30% reduction in total tumour measurement
SD	<30% reduction and <20% increase in the total tumour measurement
PD	>20% increase in the total tumour measurement or the appearance of one or more new lesions

Table 4.3: Modified RECIST for Visual Assessment of Patients with Mesothelioma

Response to therapy can be defined as complete response (CR), partial response (PR), stable disease (SD) or progressive disease (PD). Confirmed response required a repeat observation on two occasions, four weeks apart. (Modified from Byrne and Nowak, 2004).

4.3.2) SUV, SUL and SUV_{BSA}

SUV was the used quantification unit for response measures. However, SUL/SUV_{LBM} and SUV_{BSA} were also investigated to determine if there was a significant difference between the three values. A correction for plasma glucose was not used in any of the methods of quantification due to studies showing that applying a correction for glucose does not affect the reliability of SUV measures or improve the accuracy of SUV as a measure of glucose metabolism (Diederichs *et al*., 1998; Hallett *et al*., 2001). Response parameters will primarily use SUV but segmentation and response parameters were also investigated using SUL and SUV_{BSA} to assess whether there were significant differences depending on which measure of total distribution volume for quantification was used.

4.3.3) SUV_{max}

SUV_{max} was taken as the maximum SUV within the entire TV on the HERMES Hybrid Viewer software by the two readers and on PETTRA software within the segmented TV, the two being identical for all values.

4.3.4) Tumour Volume Segmentation

Semi-automated segmentation of TV was conducted with the aid of a consultant physician who decided which areas of abnormal uptake in the PET image represented tumour, as opposed to physiological uptake. For simplicity, a region growing method with a fixed 2.5 SUV threshold using 6-voxel connectivity was originally used to segment the disease in each of the 28 whole body PET images. While there are issues using a fixed SUV to segment disease, it has been successful in other studies investigating TLG in patients with mesothelioma (Veit-Haibach *et al*., 2010; Schaefer *et al*., 2012). It has the advantage that it is simple, objective, well-defined, computationally efficient and less likely to produce unpredictable results, in comparison to other

methods. While it may result in some segmentations spilling into areas of physiological uptake more often than other methods, this can be easily rectified by restricting the segmentation or removing areas of physiological uptake from areas of segmented tumour, judged by the consultant physician. If need be, disease was restricted to a given cubic volume in the image to include just tumour rather than physiological or background uptake. Equally, there were occasions where cubic volumes were removed from the segmented disease area to remove physiological or background uptake. After this segmentation method was used initially, all the segmentation methods implemented in PETTRA (described in 2.3, p91) were used to try and segment areas of tumour.

4.3.5) Tumour Volume and Total Lesion Glycolysis

TV in a PET scan was taken as the total of all segmented areas of disease. TV was calculated as the volume of each voxel multiplied by the number of segmentation voxels, given in ml. TLG was calculated as the product of TV and SUV, given in ml*SUV.

4.3.6) Intensity Volume Histogram Parameters

All six of the IVH parameters (I_{10}, I_{90}, $I_{10\text{-}90}$, V_{10}, V_{90}, $V_{10\text{-}90}$), as defined in PETTRA in 2.4.6 (p121), were investigated as potential measures of response.

4.4) Results for Segmentation of Disease in Mesothelioma Patients

4.4.1) Fixed 2.5 SUV Segmentation of Disease in Mesothelioma Patients

Disease was present in all 28 PET images of the 14 datasets. The fixed 2.5 SUV threshold region growing algorithm successfully segmented disease in all 28 images. In 21 of the 28 images, no restrictions on segmentation in PETTRA were needed to stop the segmentation of disease 'spilling' into areas of physiological uptake or the background in the image. In the remaining

seven images, nine areas of physiological uptake interfered with segmentation of disease (Table 4.4). Five of these were the heart and were removed from the segmented disease using the methodology described in 2.3.9 (Figure 2.25, p110), the one exception being the pre-therapy image in dataset 10 where disease was restricted from entering the heart using a similar methodology. All segmentations, including those where physiological uptake was removed, were viewed by a consultant who deemed the segmentation to capture diseased areas only and not areas of physiological uptake. Over the 28 images, there were a total of 290 segmented areas of disease with a mean of 10 for each image (range, 2-25). The mean TV was 2593ml (range, 7-1786ml).

While the number of areas of physiological uptake included in segmented disease was not huge, in an ideal scenario segmentation would not need the removal of unwanted uptake. Segmentation of disease in PET scans is often a more problematic task in real clinical images than in phantoms, simulated images or images where histological samples can be taken, which are often used to test segmentation algorithms. Images with high background uptake and more heterogeneous lesions are also often more difficult to segment. The pre-therapy scan for patient 6 in this dataset is a good example of the problems that can arise in segmenting PET lesions (Figure 4.2), particularly when using a fixed 2.5 SUV threshold which does not take into account background uptake. In this example, the liver and bowel have been included in the segmentation of disease and have had to be removed. Many of the higher fixed threshold segmentation methods and those that used background uptake to determine a threshold, segmented disease in the pre-therapy scan in dataset 6 (Figure 4.2) without including the liver and bowel.

Dataset	Scan	Physiological Uptake	Solution	VOIs Needed
2	Pre	Heart	Removed from segmented area	1
4	Pre	Heart	Removed from segmented area	1
6	Pre	Liver	Removed from segmented area	5
		Spleen	Removed from segmented area	1
		Bladder/Bowel	Removed from segmented area	1
	Post	Liver	Removed from segmented area	1
10	Pre	Heart	Disease restricted from heart	1
	Post	Heart	Removed from segmented area	1
13	Post	Heart	Removed from segmented area	1

Table 4.4: Physiological Uptake Segmented using a Fixed 2.5 SUV Threshold

Over the dataset, nine areas of physiological uptake were segmented with disease using a fixed 2.5 SUV region growing segmentation algorithm. Physiological uptake was either removed from the segmented area of disease or restricted to a particular volume so it did not 'spill' into areas of physiological uptake. VOIs Needed refers to the number of cubic VOIs that were needed to remove the physiological uptake.

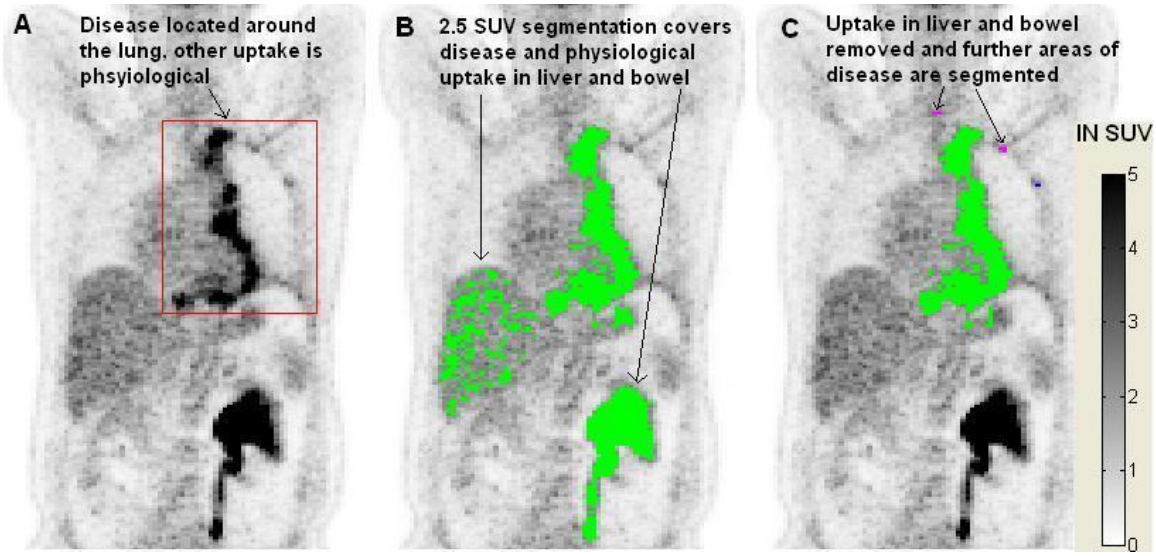


Figure 4.2: Removed Areas of Physiological Uptake for 2.5 SUV Segmentation

Coronal view of a pre-therapy PET image for dataset 6 in which both disease and physiological uptake is present (A). A fixed 2.5 SUV threshold segments both disease and background uptake in the liver and bowel (B). This is removed using restricted segmentation, before other disease is segmented (C).

4.4.2) Segmentation of Mesothelioma Patients using Different Voxel Connectivity

All segmentations used the region growing algorithm to segment disease searching for 6-connected voxels, however, searching for 18-connected or 26-connected voxels may have a significant impact on segmented TV. To test this, a fixed 2.5 SUV threshold segmentation was performed searching for 18-connected and 26-connected voxels and the results compared to 6-connected voxels. All the methods had to have the same areas of physiological uptake removed, with the removal of the liver in the post-therapy image of dataset 6 needing the co-ordinates extended to remove further physiological uptake when using 18-connected and 26-connected voxels. Using higher numbers of connected voxels resulted in needing fewer VOIs to segment the disease and increased TV and TLG (Table 4.5). The mean change in TV and TLG between pre- and post- therapy segmentations was less when using higher numbers of connected voxels. Over the 28 images, the mean difference between using 6-connected voxels and 26-connected voxels to obtain TV and TLG was 19ml (2.5%) and 54ml*SUV (2.2%) respectively. The mean difference in the change between TV and TLG between pre- and post- therapy images was 13ml (3.2%) and 36ml*SUV (2.4%). All three types of voxel connectivity correlated extremely well for both TV and TLG measurements (pcc >= 0.998, p = 0) and the change between pre- and post- therapy TV and TLG (pcc >= 0.990, p = 0). However, it should be noted that the percentage change, while still showing a strong correlation, was not as high (pcc >= 0.618, p < 0.02). Change was calculated as the post-therapy value minus the pre-therapy value and the percentage change as:

$$\% \text{ Change} = \frac{\text{Post-Therapy} - \text{Pre-Therapy Value}}{\text{Pre-Therapy Value}} * 100 \qquad [4.1]$$

Voxel Connectivity	Total VOIs	Mean		Mean Absolute Change	
		TV	TLG	TV	TLG
6-connected	290	586	2593	154	840
18-connected	249	598	2626	153	831
26-connected	233	605	2647	151	827

Table 4.5: Difference between 6-, 18- and 26-Connected Voxel Segmentation Algorithms

Table shows the total number of VOIs needed to segment disease, mean TV (ml) and mean TLG (ml*SUV) for all 28 images and the mean absolute change in TV and TLG between pre- and post- therapy scans. Segmentations were done using a fixed 2.5 SUV threshold region growing algorithm for 6-connected, 18-connected and 26-connected voxels.

These results suggest that the affect of voxel connectivity used in the segmentation algorithm is not hugely significant although changes are more significant when comparing the percentage change between pre- and post- therapy images. In this scenario, the mean difference between using 6-connected and 26-connected voxels is 30%. However, after further investigation into these results it was discovered that this was heavily influenced by one dataset having a huge positive percentage change when using 6-voxel connectivity and a more modest increase when using 26-voxel connectivity (without this dataset included, differences were <10%). On this dataset, disease in the liver was initially segmented using 6-voxel connectivity on just one side of the liver whereas with 26-voxel connectivity the algorithm identifies connected voxels above the threshold throughout the liver (Figure 4.3). It could be argued that not all the disease was segmented using the 6-voxel connectivity algorithm or that the 26-voxel connectivity algorithm has included unwanted physiological uptake, when dealing with abnormal uptake in the liver. This is hard to define or to know for sure and is one of the main issues when segmenting disease using any segmentation method. In this dataset, disease on the edge of the liver is rare and not a typical area for disease to be segmented. Without this specific example, it can otherwise be stated that the voxel connectivity used in region growing segmentation for fixed algorithms does not have a significant impact on TV or TLG.

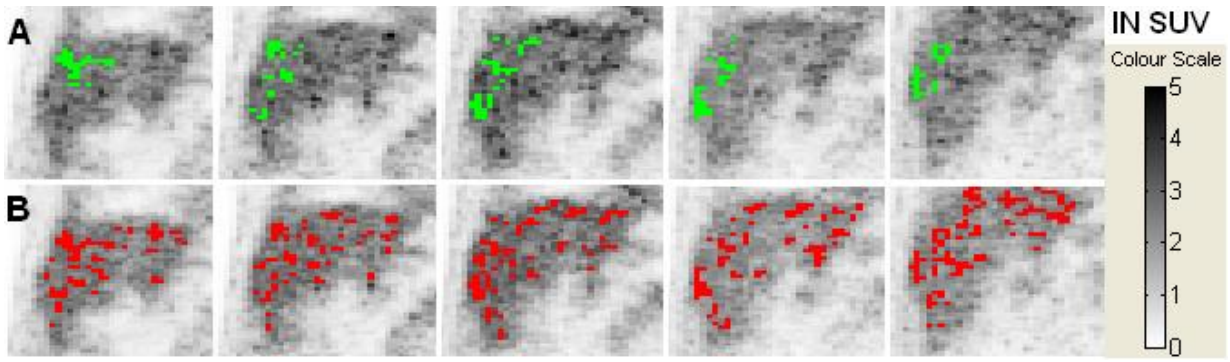


Figure 4.3: Differences in Segmentation using 6-Voxel and 26-Voxel Connectivity

Segmented disease in the liver on the pre-therapy image in dataset 8 over five connected coronal slices using both (A) 6-voxel and (B) 26-voxel connectivity for a fixed 2.5 SUV threshold region growing algorithm. While the 6-connected voxel segmentation does not spread throughout the liver, the 26-connected voxel segmentation does. This could be prevented by restricting the segmentation but for this region of disease it is difficult to determine which areas of uptake are disease and which are physiological uptake so no restrictions were performed during the segmentation process.

4.4.3) Segmentation of Mesothelioma Patients using Different Fixed Thresholds

The fixed 2.5 SUV threshold was chosen because of its use in literature (Veit-Haibach *et al*., 2010; Schaefer *et al*., 2012). However, different thresholds and variations in total distribution volume used for quantification of SUV, including SUL and SUV_{BSA}, are worthy of investigation with regard to segmentation as they may prove to be better thresholds for detecting disease. Values of 2.2, 2.4, 2.6, 2.8 and 3 SUV were used for fixed threshold segmentation, chosen at logical increments/decrements from each other. A SUV of 2 was also considered but segmentation included physiological uptake too easily on numerous occasions. For SUL, the same thresholds were used as only a small reduction in values is common from SUV, while for SUV_{BSA}, values are much lower when compared to SUV so thresholds of 0.5, 0.6 and 0.7 were used following experimentation to find suitable thresholds.

Table 4.6 shows the results of these segmentations, starting with the default fixed SUV 2.5 threshold. The higher the SUV threshold the more VOIs are used to segment all the disease in the image using region growing segmentation, as segmentation is more likely to stop between disease areas because there are less voxels connected by the higher threshold. The number of restricted VOIs used to aid segmentation is reversely proportional to the threshold value with higher thresholds meaning less restricted areas are needed. Using SUL, opposed to SUV or SUV_{BSA}, reduced the number of restricted areas needed. This is partially due to the fact that the intensity values for SUL are lower than SUV and so the same group of thresholds are effectively higher for SUL. However, to only need three restricted areas at a fixed 2.2 SUL threshold suggests that for ease of use it could be beneficial.

TV and TLG are smaller when higher thresholds are used as fewer voxels are included in segmentations as there are fewer voxels above the threshold. Mean TLG varies significantly when using SUV, SUL and SUV_{BSA} as the intensity values are greatly affected by the distribution volume used for quantification. However, mean TV can give an idea at what threshold similar volumes for SUV, SUL and SUV_{BSA} are obtained with TV relatively similar when using thresholds of 0.7 SUV_{BSA}, 2.2 SUL and 2.6-2.8 SUV. The absolute change (i.e. decrease or increase, both taken as a positive value) in TV and TLG is larger at lower thresholds as there is more disease included in segmentations so there is more likely to be greater changes in the quantity of disease between pre- and post- therapy scans. The mean absolute percentage change between TV and TLG is larger at higher thresholds, as they segment smaller regions of disease which are likely to produce a greater percentage change than thresholds which produce larger segmented areas of disease.

Fixed Threshold	Total VOIs	Restricted Areas	Mean TV	Mean TLG	Abs Change		Abs % Change	
					TV	TLG	TV	TLG
2.5 SUV	290	13	586	2593	154	840	82	89
2.2 SUV	229	16	760	3008	182	872	65	70
2.4 SUV	286	16	636	2716	165	847	72	78
2.6 SUV	354	9	545	2487	142	818	64	71
2.8 SUV	435	5	481	2313	136	825	77	86
3 SUV	464	5	428	2159	133	832	85	96
2.2 SUL	278	3	504	1930	141	650	104	116
2.4 SUL	337	3	437	1776	133	662	150	169
2.5 SUL	437	3	408	1705	130	664	148	167
2.6 SUL	388	3	383	1640	129	666	174	197
2.8 SUL	410	3	338	1518	128	668	236	264
3 SUL	436	3	301	1411	126	660	452	492
0.5 SUV_{BSA}	186	24	996	912	260	260	95	93
0.6 SUV_{BSA}	332	11	677	736	177	230	121	126
0.7 SUV_{BSA}	477	5	517	632	144	212	103	114

Table 4.6: Difference between Fixed SUV, SUL and SUV_{BSA} Threshold Segmentation

Table shows the total number of VOIs needed to segment disease, number of times restricted areas were used to stop segmentation, the mean TV (ml) and TLG (ml*SUV) for all 28 images and mean absolute change and absolute percentage change between pre- and post- therapy scans for all 14 datasets. All segmentation algorithms used threshold region growing algorithms using 6-connected voxels.

The visual difference between these segmentation methods is difficult to show, as 3-D VOIs are difficult to illustrate, however, a comparison of different SUV thresholds on the main regions of disease for three pre-therapy images is displayed in Figure 4.4. The effects of different thresholds can be seen in dataset 1 and dataset 12. In dataset 1, an area of disease uptake near the top of the lung is not included at a threshold of SUV 2.5 but is included using an SUV 2.4 while in dataset 12 a similar scenario occurs between SUV 2.5 and 2.6. The missing areas of disease can be segmented using an additional VOI, however, in an ideal scenario segmentation should include these areas without spilling into physiological uptake.

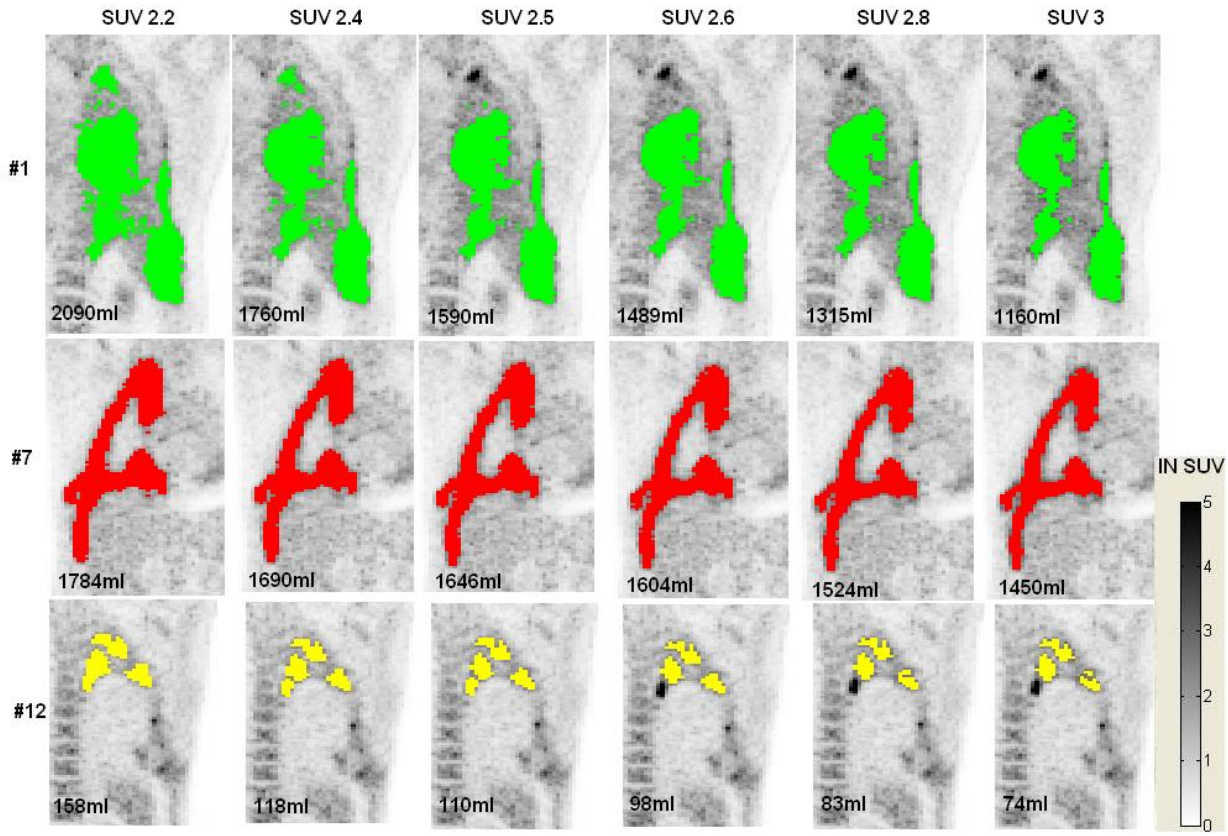


Figure 4.4: Segmentation using Different Fixed SUV Thresholds

Segmentation of the primary disease region on pre-therapy scans for datasets 1, 7 and 12 using a fixed region growing algorithm for 2.2, 2.4, 2.5, 2.6, 2.8 and 3 SUV. TV is shown in the bottom left corner of each image. Images are of the coronal slice in which the SUV_{max} of the disease is present.

Dataset 7 shows the opposite scenario, all SUV thresholds segment the same main area of disease, which is less heterogeneous and more connected making this an easier task, with lower thresholds segmenting more voxels on the edge of the tumour as would be expected. Figure 4.5 shows the same datasets for SUL and SUV_{BSA} at various thresholds. Different thresholds using SUL seem to make less of a difference to segmentation in the three datasets, as they appear similar with TV changing moderately, as would be expected. For the three SUV_{BSA} thresholds tested, variation was apparent but there is more of a gap, proportionately, between thresholds.

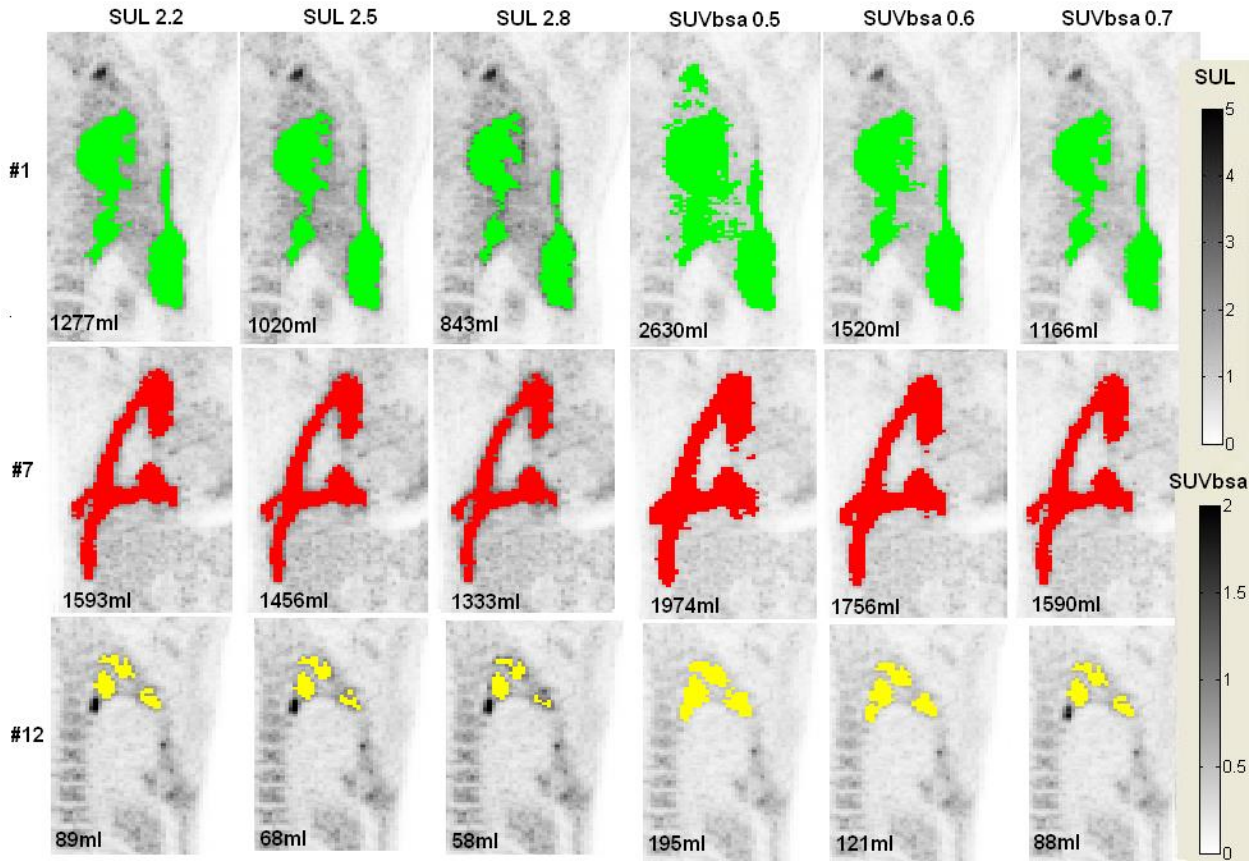


Figure 4.5: Segmentation using Different Fixed SUL and SUV_{BSA} Thresholds

Segmentation of the primary disease region on pre-therapy scans for datasets 1, 7 and 12 using a fixed region growing algorithm for 2.2, 2.5, 2.8 SUL and 0.5, 0.6, 0.7 SUV_{BSA}. TV is shown in the bottom left corner of each image. Images are of the coronal slice in which the SUV_{max} of the disease is present. Images using SUL are shown with the default colour range of 0-5 SUL, where as images using SUV_{BSA} are shown using a colour range of 0-2.

The correlation of $SUV_{max,}$ TV and TLG parameters between segmentation using 2.5 SUV and those with different thresholds and normalisation methods is strong (pcc >= 0.850, $p < 0.001$ in all but one case, Table 4.7). Even when investigating the change between pre- and post- therapy scans, correlations are still significant showing that even though there is an impact when using different fixed thresholds, it is unlikely to have an effect on the prediction of response. The one instance where correlation was weak was between the change in TV using 2.5 SUV and 0.5 SUV_{BSA} thresholds (pcc = 0.283, $p < 0.33$). This can be explained by the significant amount of extra disease, possibly some of which could be physiological uptake, which may have been included due to the low threshold (once again, the segmentation of the disease in the liver is likely to be part of the reason for this in a similar way to the difference between 6-connected voxel segmentation and 26-connected voxel segmentation, Figure 4.3). This may have had enough of an influence to cause significant differences in the change between TV in comparison to using a 2.5 SUV threshold, explaining its lower correlation coefficient. The TLG correlation with a fixed 2.5 SUV segmentation is also lower than others for this method. However, it is still high in comparison to TV correlations, because even with increased TV, the SUV intensity within the TV is likely to remain more stable as the highest intensity regions will always be segmented no matter what the segmentation method.

Fixed Threshold	Over all 28 images			Change			% Change		
	Max	TV	TLG	Max	TV	TLG	Max	TV	TLG
2.5 SUV	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2.2 SUV	1.000	0.992	0.997	1.000	0.850	0.984	1.000	0.897	0.924
2.4 SUV	1.000	0.999	1.000	1.000	0.983	0.998	1.000	0.995	0.995
2.6 SUV	1.000	0.999	1.000	1.000	0.980	0.997	1.000	0.902	0.917
2.8 SUV	1.000	0.996	0.998	1.000	0.963	0.996	1.000	0.963	0.973
3 SUV	1.000	0.990	0.996	1.000	0.943	0.993	1.000	0.902	0.918
2.2 SUL	0.993	0.993	0.995	0.994	0.961	0.991	0.996	0.977	0.981
2.4 SUL	0.993	0.989	0.992	0.994	0.953	0.992	0.996	0.963	0.966
2.5 SUL	0.993	0.985	0.990	0.994	0.946	0.991	0.996	0.964	0.968
2.6 SUL	0.993	0.981	0.988	0.994	0.938	0.991	0.996	0.924	0.933
2.8 SUL	0.993	0.971	0.984	0.994	0.924	0.989	0.996	0.875	0.887
3 SUL	0.993	0.960	0.978	0.994	0.914	0.988	0.996	0.856	0.875
0.5 SUV_{BSA}	0.981	0.967	0.979	0.981	0.283	0.881	0.995	0.582	0.678
0.6 SUV_{BSA}	0.981	0.988	0.985	0.981	0.859	0.967	0.995	0.958	0.960
0.7 SUV_{BSA}	0.981	0.983	0.982	0.981	0.941	0.977	0.995	0.980	0.986

Table 4.7: Correlation between Fixed SUV, SUL and SUV_{BSA} Threshold Segmentations

Pearson correlation coefficients (pccs) are shown for the correlation of various fixed thresholds of SUV, SUL and SUV_{BSA} with a 2.5 SUV threshold for SUV_{max}, TV and TLG. Pccs were calculated for SUV_{max}, TV and TLG over the 28 pre- and post- therapy images and for the change and percentage change between the 14 pre- and post- therapy scans. For $p < 0.01$, $pcc > 0.479$ over the 28 images and for $p < 0.01$, $pcc > 0.662$ for changes and % change.

4.4.4) Segmentation of Mesothelioma Patients using Percentage of SUV_{max}

While segmentation using a percentage of SUV_{max} has been used successfully for segmenting disease in many recent cancer studies (Chen *et al*., 2012; Chung *et al*., 2012; Dibble *et al*., 2012; Im *et al*., 2012; Lim *et al*., 2012), it is not the best technique to use when segmenting mesothelioma patients. Figure 4.6 shows a comparison of segmentations using a given percentage of SUV_{max} over three datasets. Using 40-70% of SUV_{max} gives thresholds which are too high for segmenting large regions of disease accurately, resulting in severe under segmentation of the main region of disease in the PET image. For example, using 40-70% of SUV_{max} segmentation gives TVs of <23ml for the main region of disease in dataset 1 compared to 1590ml using a fixed

2.5 SUV threshold segmentation. To get a TV comparable with 1590ml, deemed suitable segmentation of the region of disease by a consultant physician, a threshold of 13%-14% of SUV_{max} is needed (TV: 1749ml using 13% and 1496ml using 14%).

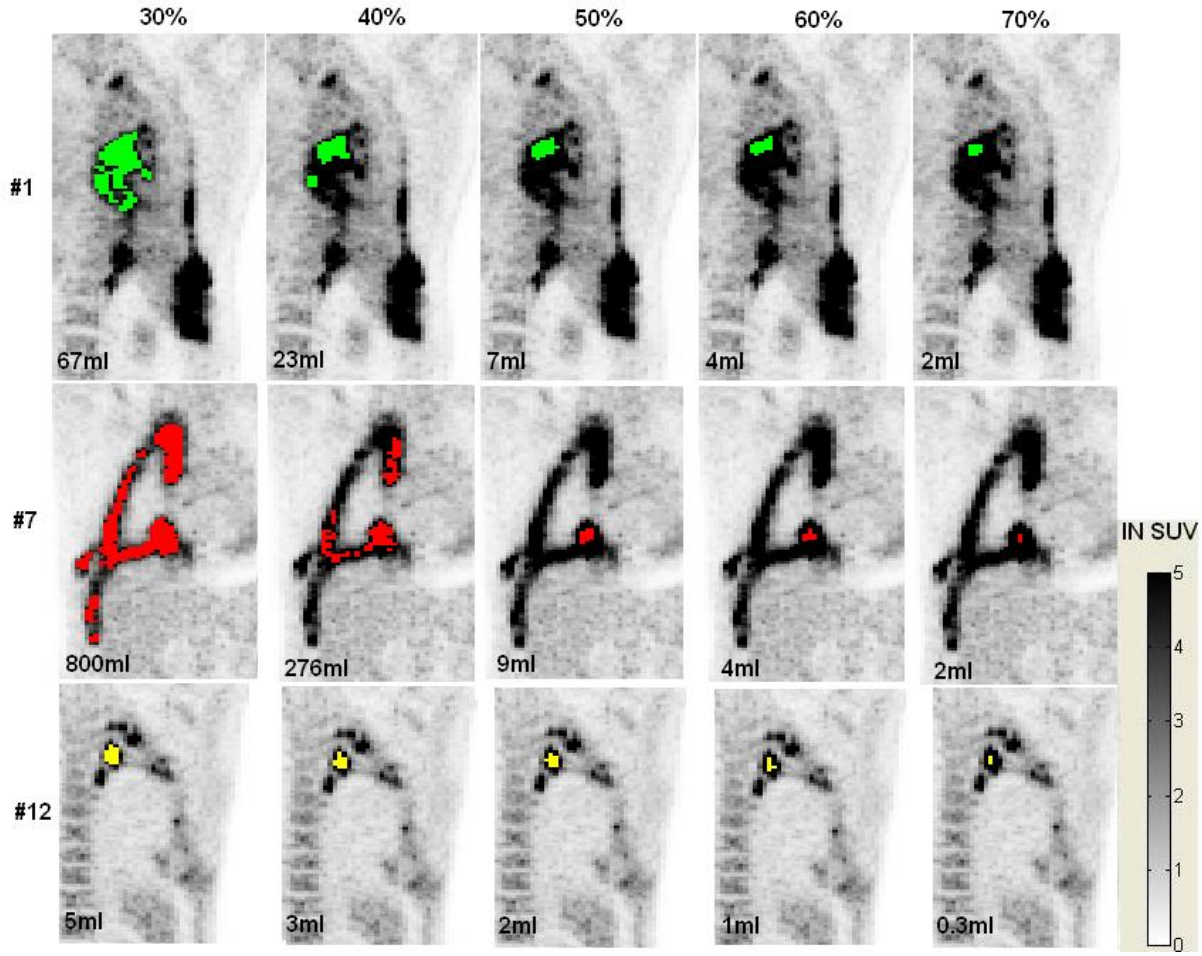


Figure 4.6: Segmentation using Percentage of SUV_{max}

Segmentation of the primary disease region on pre-therapy scans for datasets 1, 7 and 12 using fixed region growing algorithm for 30%, 40%, 50%, 60% and 70% of SUV_{max}. TV is shown in the bottom left corner of each image. Images are of the coronal slice in which the SUV_{max} of the disease is present.

While this might work for large regions of disease, using this sort of threshold on smaller regions results in very low thresholds which then severely over-segment disease and include huge areas

of background and physiological uptake. For example, for dataset 1, using the same 14% SUV_{max} threshold on a smaller lesion in the patient, not connected to the rest of the disease, produces a threshold of 0.8 SUV and a large TV of 16,341ml which segments almost the entire patient including physiological uptake in the brain, liver, bowel and bladder (Figure 4.7). In comparison, segmentation using a 2.5 SUV threshold produces a TV of just 2ml. While segmentation using a fixed 2.5 SUV threshold is not perfect, it has been shown to produce suitable results as assessed by a consultant physician and, in comparison, using a percentage of SUV_{max} is unreliable and prone to vastly over- or under- segment disease (Figure 4.6, Figure 4.7). Therefore, segmentation using percentage of the SUV_{max} was not investigated further.

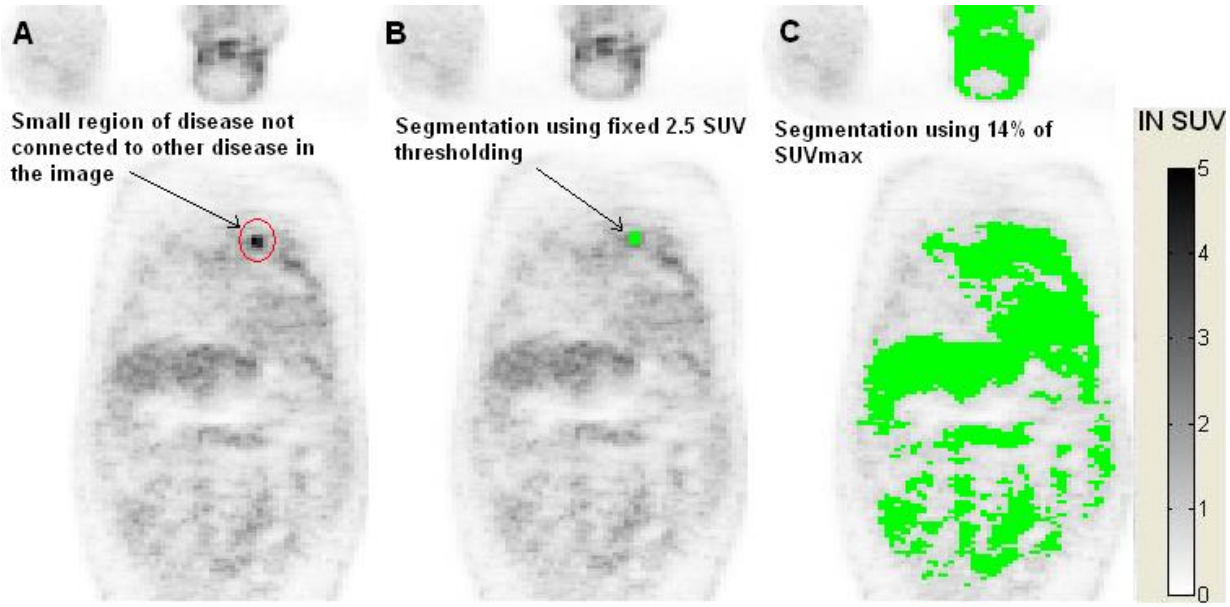


Figure 4.7: Issues with Segmentation using a Percentage of SUV_{max}

Segmentation of the primary disease region on pre-therapy scan for dataset 1 needs a suitable threshold of around 14% of SUV_{max}. When this is applied to a smaller disease lesion (A), segmented with at least some degree of visual success using a fixed 2.5 SUV threshold (B), it clearly spills into areas of physiological uptake including brain, liver and bowel with a threshold of just 0.8 SUV (C).

This does not mean that using a percentage of SUV_{max} has no place in PET tumour segmentation. For many cancers where lesions are typically homogeneous and of similar small sizes, it has been shown to work competently enough (Chen *et al*., 2012; Chung *et al*., 2012; Dibble *et al*., 2012; Im *et al*., 2012; Lim *et al*., 2012). However, for mesothelioma, where there is more heterogeneous uptake and both small and large regions of disease, it has been found to be inadequate at reliably segmenting TV.

4.4.5) Segmentation of Mesothelioma Patients using SUV_{mean}

Segmentation using SUV_{mean} can be done in the form of a threshold based method which relates SUV_{mean} to a threshold using two fixed constants (Black *et al*., 2004), and an adaptive region growing algorithm using a fixed percentage of SUV_{mean} (Green *et al*., 2008). Both are implemented in PETTRA as described in 2.3.5 (p99). In both studies, the segmentation methods were tested on phantom studies so experimentation was done on the mesothelioma dataset to test their viability before pursuing with these methods. Modifications were made to the constants and percentages used where necessary. Figure 4.8 shows the performance of the adaptive SUV_{mean} based segmentation method for three pre-therapy images on the main region of disease using thresholds of 50%, 60%, 65%, 70% and 85% of SUV_{mean} using the adaptive algorithm.

The problems with using the adaptive mean algorithm using a percentage of SUV_{mean} are highlighted when viewing Figure 4.8. The 85% threshold used on phantom studies is too high on clinical data of mesothelioma patients, under-segmenting TV while a threshold of 50% appears too low, as can be seen on dataset 1 where the segmentation has ‘escaped’ from the tumour (region of local maxima), ending up with a very low threshold which then floods into other regions of the PET scan so the maximum voxel is in a different PET structure, in this case the bladder (region of global maxima). This issue is illustrated further in Figure 4.9. Despite this, the adaptive SUV_{mean} algorithm was used on the dataset using 60%, 65% and 70% thresholds and a

70% threshold was found to work successfully enough to segment TV in 26/28 images with the percentage level increased to 80% on one dataset (dataset 14) and 90% on another (dataset 8) to ensure valid segmentation could take place. The changes in threshold percentages were used if the seed voxel fell outside the final VOI when using a 70% threshold, for any segmentation where the seed voxel falls outside the final VOI is rejected. The potential issues with this method are dealt with in its later incarnation, GRAB, which uses background uptake (Boucek *et al*., 2008).

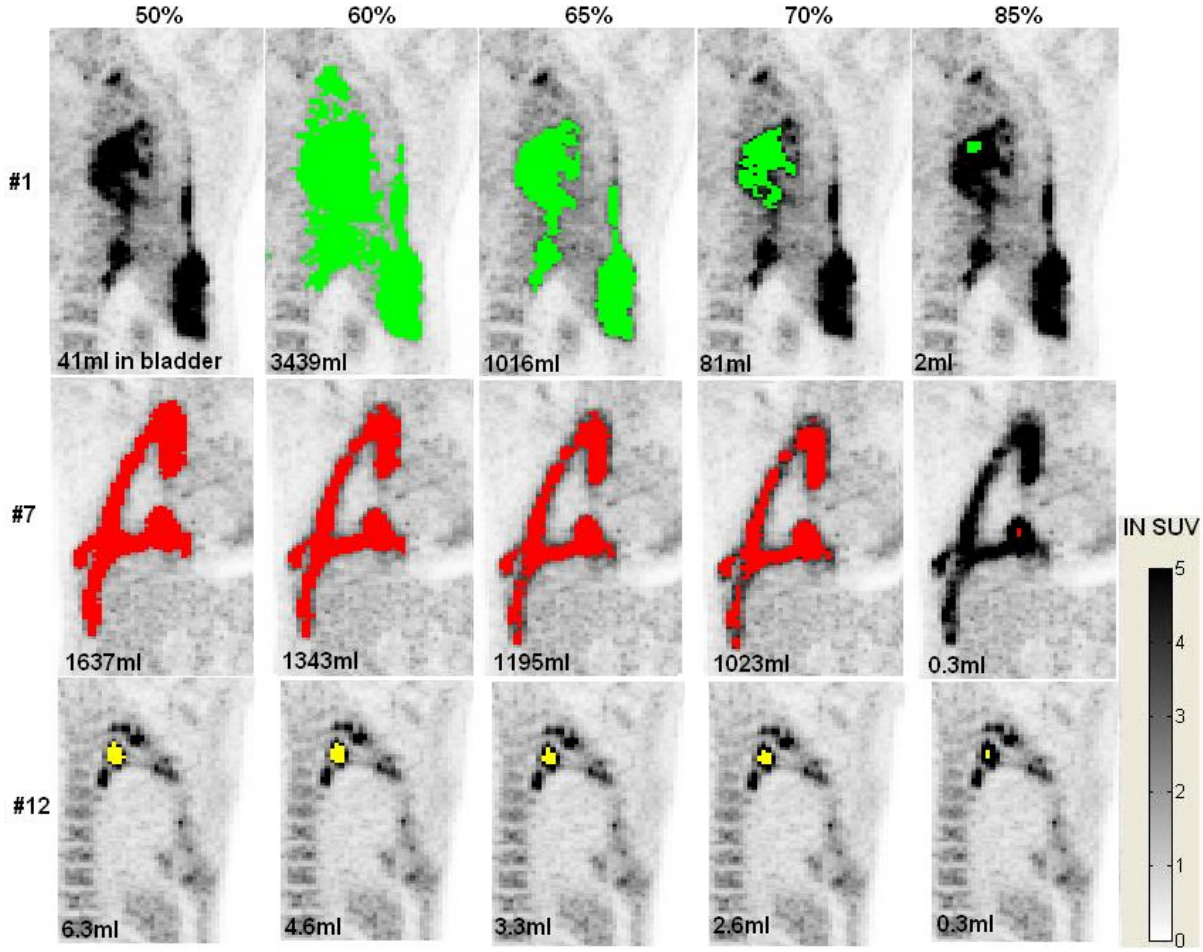


Figure 4.8: Segmentation using Adaptive SUV_{mean}

Segmentation of the primary disease region on pre-therapy scans for datasets 1, 7 and 12 using adaptive SUV_{mean} segmentation using thresholds of 50%, 60%, 65%, 70% and 85% of SUV_{mean}. TV is shown in the bottom left corner of each image. Images are of the coronal slice in which the SUV_{max} of the disease is present. For dataset 1 at a 50% threshold the segmentation has flooded into the bladder and segmented the bladder instead of the main region of disease.

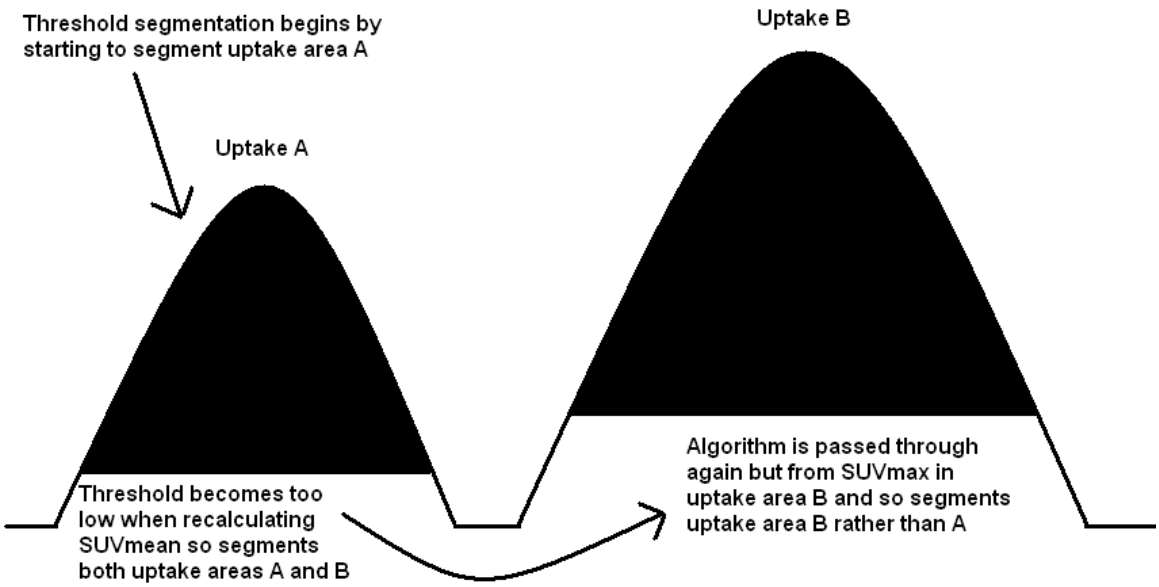


Figure 4.9: Issue of Local Maxima in Segmentation using Adaptive SUV_{mean}

If the percentage used in the adaptive SUV_{mean} segmentation is too low the threshold will get to a point where it is low enough to 'flood' into other regions of the image and the SUV_{mean} will take into account other uptake areas and end up segmenting another PET structure.

Similar testing was conducted on the SUV_{mean} method using a threshold calculation involving two fixed constants (Black *et al*., 2004). The given constants of 0.307 and 0.588 visually appeared to be too low and included physiological uptake while those with higher constant values seemed to provide better segmentation (Figure 4.10). Different calculations using alternative constant values were decided by logical increments in values to provide higher threshold values. All four modified calculations used in Figure 4.10 were used in segmenting the dataset although the calculation using the lowest constant values ($0.4 * SUV_{mean} + 0.6$) was found to give thresholds which were too low on some datasets where it was impossible to restrict disease from the background or remove background regions. For this reason, this variation of the calculation was not used for further analysis.

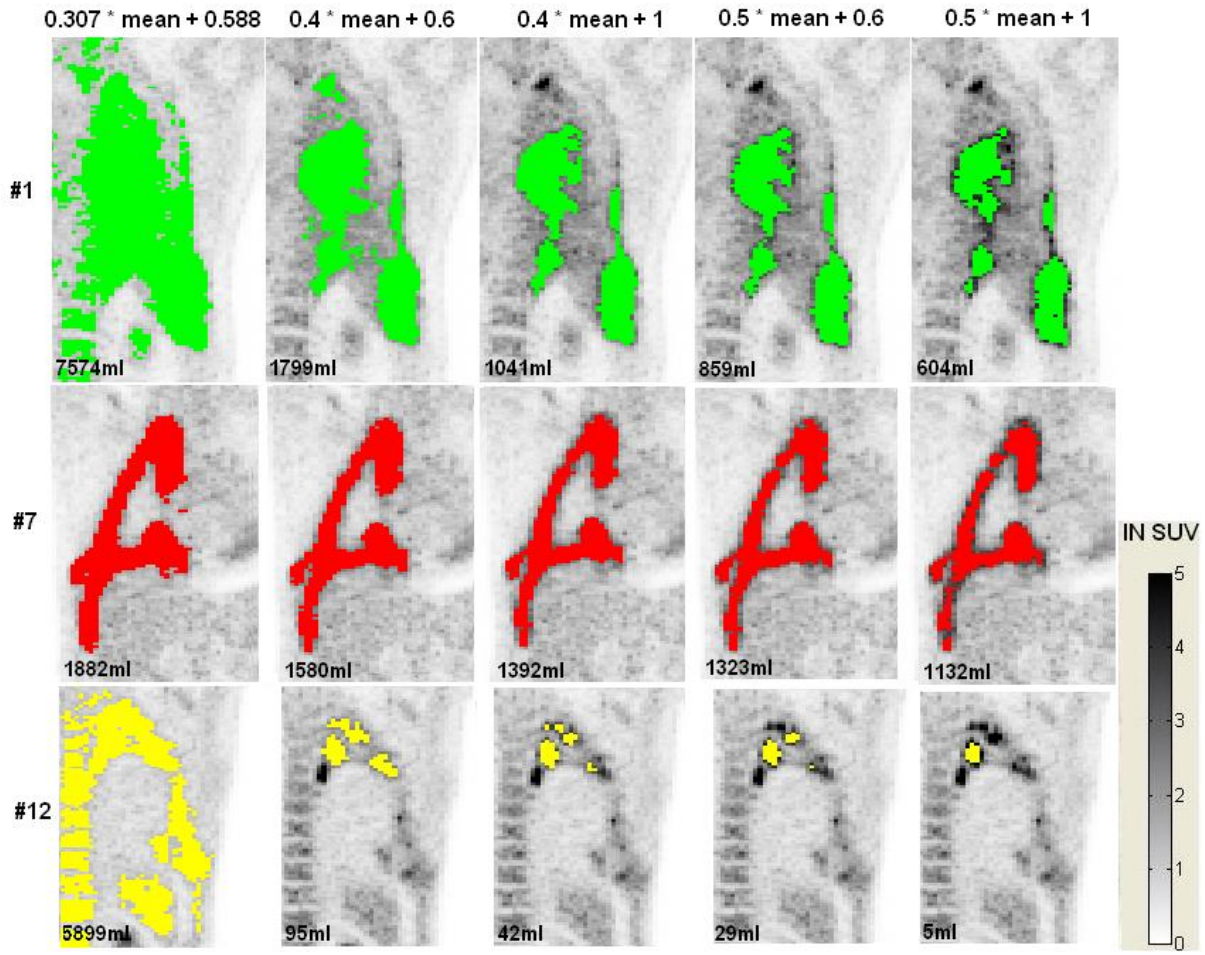


Figure 4.10: Segmentation using SUV_{mean} Threshold Calculation Method

Segmentation of the primary disease region on pre-therapy scans for datasets 1, 7 and 12 using SUV_{mean} calculation method using various values for the constants (Black *et al.*, 2004). Calculations used were 0.307 * SUV_{mean} + 0.588, 0.4 * SUV_{mean} + 0.6, 0.4 * SUV_{mean} + 1, 0.5 * SUV_{mean} + 0.6 and 0.5 * SUV_{mean} + 1. TV is shown in the bottom left corner of each image. Images are of the coronal slice in which the SUV_{max} of the disease is present.

For all threshold calculations using SUV_{mean}, it is assumed that the SUV_{mean} is known and, therefore, determining the value used for SUV_{mean} is problematic. In PETTRA, implementation is achieved by using a starting SUV_{mean} obtained from a fixed 2.5 SUV threshold segmentation and then recalculating the threshold until the TV does not change significantly, as described in 2.3.5

(p99). It was important to ensure that initial segmentation did not heavily influence the method. Therefore, other fixed thresholds from 2.2 SUV to 4 SUV were used on the main regions of disease on pre-therapy images in the dataset to see if there would be any significant changes in the final fixed threshold using the SUV_{mean} calculation method (Black *et al*., 2004). Differences were found to be <1% on all occasions so the starting value of SUV was considered irrelevant and kept at 2.5 SUV.

The SUV_{mean} calculation algorithm stops when the change in TV is <1% or after 10 iterations of re-calculating the threshold value based on the new SUV_{mean} (as described in 2.3.5, p99). In testing the SUV_{mean} calculation method, it was found that the algorithm often gets 'stuck' between two different PET structures. For example, when using the calculation of 0.5 * SUV_{mean} + 1 for a small tumour region on dataset 1, the algorithm reaches a point where the threshold is 4 SUV, giving a 140 voxel TV with a SUV_{mean} of 5.1. Using the calculation, the new threshold becomes 3.55 SUV, giving an 8840 voxel TV with a SUV_{mean} of 6. The recalculation of the threshold sets it back to 3 and this loop continues. One of the thresholds needs to be used for the segmentation and the algorithm has been modified to take the smaller volume/higher threshold to stop separate segmentations overlapping each other. In line with this change, the number of iterations of the algorithm performed was increased to 20 rather 10 to ensure that if convergence was possible it would take place. It was found that convergence to a volume change of <1% sometimes did not occur until a higher number of iterations had been performed. No other changes were made to the algorithm and the three suggested calculation variations (0.4 * SUV_{mean} + 1, 0.5 * SUV_{mean} + 0.6 and 0.5 * SUV_{mean} + 1) were used to segment all 14 datasets of mesothelioma patients successfully. These same changes used in this segmentation method were also applied to similar methods which used this methodology of obtained the SUV_{mean}, mainly those by Nestle *et al*. (Nestle *et al*., 2005; Nestle *et al*., 2007).

4.4.6) Background Uptake for Segmentation

A major factor in segmentations using background uptake is the methodology used to obtain the background value (and S.D., if needed). While automated or semi-automated methods are possible and will eliminate or limit variation due to observer variability, there is no standard method used in clinical practice. Therefore, a manual methodology was used because of its easy of use and the opportunity to assess the levels of variation in background uptake. One observer placed eight different sized VOIs (listed in Table 4.8) over three background regions (middle of the liver, top of liver and middle of the mediastinum) for all 28 PET images. VOIs ranged from ~2ml to >200ml. Seven were fixed VOIs while another VOI was taken as a user defined volume over the whole of the liver.

The liver and mediastinum were chosen as the regions of background uptake due to recommendations of the PERCIST criteria (Wahl *et al*., 2009). The one observer had some experience of viewing PET images and had no issues identifying the liver in all 28 images but found it more difficult to identify the mediastinum, particular a region which could be defined as the 'middle'. All 22 VOIs (seven fixed VOIs over the three background regions and the user defined liver region) were used over the 14 datasets, however, if the observer believed that the VOI was too large for the region or would include areas of disease then it was not included in the analysis. While the presence of disease in and around the lung made this a particularly problem in the mediastinum, in the liver there were only three datasets where disease was a problem. In two of these, disease was only present on the outer areas so background regions could be obtained with care, although on dataset 8, where disease was present in the liver, it was more likely some background regions contained areas of disease.

VOIs used for 27 images with voxel sizes of 4.69mm x 4.69mm x 3.27mm (97.80mm)								
VOI	mm			mm^3	Voxels			
	X	Y	Z	Volume	X	Y	Z	Volume
2.6 ml VOI	16.41	16.41	9.81	2,642	3	3	3	27
12 ml VOI	27.35	27.35	16.35	12,230	5	5	5	125
22 ml VOI	27.35	27.35	29.43	22,014	5	5	9	225
34 ml VOI	38.29	38.29	22.89	33,560	7	7	7	343
53 ml VOI	38.29	38.29	35.97	52,736	7	7	11	539
119 ml VOI	49.23	49.23	49.05	118,877	9	9	15	1215
201 ml VOI	60.17	60.17	55.59	201,260	11	11	17	2057
User Defined (Mean)	111.04	101.33	103.17	801,731	20	19	32	8198

VOIs used for image with voxel sizes 4.69mm x 4.69mm x 3.27mm (71.85mm)								
VOI	mm			mm^3	Voxels			
	X	Y	Z	Volume	X	Y	Z	Volume
2 ml VOI	14.07	14.07	9.81	1,942	3	3	3	27
13 ml VOI	23.45	23.45	22.89	12,587	5	5	7	175
25 ml VOI	32.83	32.83	22.89	24,671	7	7	7	343
32 ml VOI	32.83	32.83	29.43	31,720	7	7	9	441
52 ml VOI	42.21	42.21	29.43	52,435	9	9	9	729
113 ml VOI	51.59	51.59	42.51	113,142	11	11	13	1573
210 ml VOI	70.35	70.35	42.51	210,387	15	15	13	2925
User Defined (Mean)	100.84	96.15	73.58	474,935	22	21	23	8189

Table 4.8: Different Sizes of VOI used for Obtaining Background Uptake

Table shows the eight different VOIs used over the 28 images for obtaining background. All seven fixed VOIs were applied to all three areas of background uptake (top of the liver, middle of the liver and the mediastinum) while the final user defined VOI was applied to the liver only. VOIs are given in voxels and mms for each plane and for the 3-D volume.

While disease should not be present in areas of background uptake, it was felt that a background region in the liver had to be used for dataset 8, as the background mean in the mediastinum was found to be too low to segment the disease in the liver accurately using many different segmentation methods. Therefore, background uptake was taken from the liver although the volume used was careful not to include the most intense regions of disease uptake. Otherwise, there were no issues with disease being included in background regions throughout the dataset.

Intra-observer variability was investigated by the one observer identifying each size VOI over the three regions twice and the differences between the obtained values (SUV_{max}, SUV_{mean}, S.D. and SUV_{mean} + 2 S.D, the calculation used to determine the segmentation threshold in PERCIST) were calculated. Table 4.9 shows the mean absolute and percentage differences for each VOI size and region and over all VOIs.

VOI	Mean Difference (% in Brackets)				
	Used	SUV_{max}	SUV_{mean}	S.D.	Mean + 2 S.D.
Mediastinum		**0.232 (2.35)**	**0.115 (1.88)**	**0.042 (3.55)**	**0.156 (1.83)**
2.5 ml VOI	13	0.330 (3.65)	0.194 (3.28)	0.081 (6.63)	0.310 (3.67)
13 ml VOI	12	0.298 (2.86)	0.122 (1.90)	0.037 (3.22)	0.176 (2.01)
22 ml VOI	11	0.127 (1.30)	0.076 (1.28)	0.035 (3.01)	0.102 (1.24)
34 ml VOI	11	0.238 (2.27)	0.091 (1.45)	0.030 (2.54)	0.101 (1.17)
53 ml VOI	8	0.154 (1.45)	0.078 (1.26)	0.027 (2.25)	0.076 (0.85)
119 ml VOI	4	0.153 (1.45)	0.080 (1.29)	0.018 (1.43)	0.060 (0.72)
Top of the Liver		**0.183 (1.55)**	**0.058 (0.78)**	**0.028 (2.32)**	**0.081 (0.83)**
2.5 ml VOI	14	0.307 (3.07)	0.130 (1.71)	0.069 (6.21)	0.215 (2.25)
13 ml VOI	14	0.180 (1.45)	0.059 (0.79)	0.024 (1.92)	0.085 (0.84)
22 ml VOI	14	0.183 (1.60)	0.061 (0.84)	0.025 (2.04)	0.070 (0.71)
34 ml VOI	14	0.138 (1.18)	0.028 (0.40)	0.019 (1.50)	0.038 (0.40)
53 ml VOI	14	0.252 (1.82)	0.046 (0.64)	0.023 (1.75)	0.059 (0.63)
119 ml VOI	13	0.092 (0.68)	0.035 (0.49)	0.012 (0.90)	0.033 (0.35)
201 ml VOI	7	0.063 (0.47)	0.031 (0.40)	0.022 (1.39)	0.038 (0.35)
Middle of the Liver		**0.120 (1.00)**	**0.047 (0.63)**	**0.018 (1.41)**	**0.055 (0.55)**
2.5 ml VOI	14	0.148 (1.43)	0.092 (1.24)	0.028 (2.44)	0.123 (1.24)
13 ml VOI	14	0.177 (1.55)	0.044 (0.60)	0.021 (1.74)	0.066 (0.68)
22 ml VOI	14	0.131 (1.06)	0.039 (0.53)	0.020 (1.60)	0.065 (0.65)
34 ml VOI	14	0.130 (1.09)	0.034 (0.47)	0.018 (1.39)	0.047 (0.48)
53 ml VOI	14	0.084 (0.73)	0.040 (0.57)	0.015 (1.21)	0.038 (0.39)
119 ml VOI	14	0.130 (0.95)	0.045 (0.61)	0.012 (0.89)	0.043 (0.42)
201 ml VOI	13	0.082 (0.56)	0.031 (0.44)	0.011 (0.82)	0.026 (0.27)
User Defined	11	0.064 (0.45)	0.046 (0.58)	0.018 (1.07)	0.019 (0.17)
ALL		**0.168 (1.50)**	**0.066 (0.97)**	**0.027 (2.22)**	**0.087 (0.94)**

Table 4.9: Intra-Observer Variability for Obtaining Background Uptake

Table shows the difference between the SUV_{max}, SUV_{mean}, S.D. and SUV_{mean} + 2 S.D (recommended threshold calculation for segmentation in PERCIST) for each VOI size over each region of background. Used = the number of datasets of the 14 the VOI was validly taken for. User Defined was a volume taken over the largest region of the liver possible, focused on the middle. % Difference is in brackets.

Overall, differences were minimal with the mean difference in SUV_{mean} just 0.066 (<1%) and the difference in SUV_{mean} + 2 S.D, just 0.087 (<1%). Using larger VOIs showed less variation as the likelihood of the placement of the VOIs overlapping was greater, reducing variability. The region with the least variation was the middle of the liver, followed by the top of the liver. Given these results, it seems appropriate to use the largest VOI possible in the middle of the liver to define background uptake in terms of reducing observer variability.

The mean values over the 14 datasets for each size VOI and region of uptake were calculated and are presented in Table 4.10. The background mean over all VOIs and regions of uptake was 1.80, however, mean values in the liver were higher (1.85 in the middle and 1.92 at the top) compared to those in the mediastinum (1.53). The same trend was true for S.D. with values higher in the liver than the mediastinum. While there is no 'known' value of background to compare these to, it is more appropriate to use the higher intensity region of background uptake to ensure that the thresholds used for segmentation are high enough to include areas of disease only, therefore, once again choosing a background region in the liver (either in the middle or at the top) seems to be the most sensible option.

Table 4.11 shows that both SUV_{mean} and S.D. were relatively consistent over the different sized VOIs and, as a consequence, the formula for PERCIST threshold calculation (Mean + 2 S.D.). From the results, it can be concluded that obtaining a background region has low intra-observer variability and does not appear to alter too much depending on the size of the VOI. However, the region in which the background is taken is significant. Given these findings, it was decided that the background mean and S.D. for the segmentation methods would be based on the largest sized VOI taken in the middle of the liver. This was 201ml on 13 of the 14 datasets and 119ml on the remaining dataset.

VOI	Mean Values over 14 Datasets			
	Used	SUV_{max}	SUV_{mean}	Mean + 2 S.D.
Mediastinum		**2.48 (0.42)**	**1.53 (0.24)**	**2.13 (0.33)**
2.5 ml VOI	13	2.27 (0.37)	1.58 (0.27)	2.19 (0.33)
13 ml VOI	12	2.39 (0.39)	1.54 (0.23)	2.09 (0.31)
22 ml VOI	11	2.45 (0.48)	1.50 (0.26)	2.07 (0.37)
34 ml VOI	11	2.57 (0.40)	1.51 (0.22)	2.13 (0.35)
53 ml VOI	8	2.69 (0.49)	1.52 (0.22)	2.14 (0.35)
119 ml VOI	4	2.86 (0.39)	1.50 (0.16)	2.15 (0.25)
Top of the Liver		**3.07 (0.59)**	**1.92 (0.35)**	**2.55 (0.47)**
2.5 ml VOI	14	2.59 (0.53)	1.93 (0.38)	2.52 (0.52)
13 ml VOI	14	2.86 (0.59)	1.91 (0.37)	2.52 (0.50)
22 ml VOI	14	2.98 (0.59)	1.91 (0.36)	2.52 (0.48)
34 ml VOI	14	3.14 (0.61)	1.90 (0.37)	2.54 (0.48)
53 ml VOI	14	3.23 (0.58)	1.89 (0.35)	2.54 (0.45)
119 ml VOI	13	3.35 (0.62)	1.92 (0.32)	2.59 (0.43)
201 ml VOI	7	3.63 (0.64)	2.00 (0.25)	2.74 (0.38)
Middle of the Liver		**3.05 (0.62)**	**1.85 (0.32)**	**2.49 (0.45)**
2.5 ml VOI	14	2.50 (0.42)	1.86 (0.30)	2.42 (0.40)
13 ml VOI	14	2.85 (0.61)	1.84 (0.32)	2.45 (0.44)
22 ml VOI	14	2.91 (0.67)	1.85 (0.33)	2.46 (0.46)
34 ml VOI	14	3.02 (0.60)	1.86 (0.32)	2.48 (0.44)
53 ml VOI	14	3.06 (0.63)	1.85 (0.33)	2.48 (0.45)
119 ml VOI	14	3.20 (0.61)	1.87 (0.33)	2.51 (0.45)
201 ml VOI	13	3.41 (0.73)	1.87 (0.35)	2.56 (0.48)
User Defined	11	3.64 (0.69)	1.83 (0.29)	2.60 (0.47)
ALL		**2.93 (0.56)**	**1.80 (0.31)**	**2.43 (0.43)**

Table 4.10: Mean Values for Background Uptake for Different VOIs

Table shows the mean values for SUV_{max}, SUV_{mean}, S.D. and SUV_{mean} + 2 S.D (recommended threshold calculation for segmentation in PERCIST) for each VOI size over each region of background. Used = the number of datasets of the 14 the VOI was validly taken for. User Defined was a volume taken over the largest region of the liver possible, focused on the middle. S.D. is in brackets. For each dataset the values for the VOIs were taken as the mean of the two attempts made at taking a region by the observer.

4.4.7) Segmentation of Mesothelioma Patients using PERCIST Recommendations

Segmentation using background uptake is recommended in PERCIST (Wahl *et* al., 2009), and is implemented in PETTRA as described in 2.3.6 (p102). It calculates the threshold based on the background mean + 2 S.D., an earlier study used a similar calculation using 3 S.D. (Zasadny *et al*., 1998). Both of these will be tested on the dataset. There were no problems with segmentation

using either of these methods. The background values used for all segmentation methods and the thresholds used for PERCIST segmentation are shown in Table 4.11. The mean value for the background uptake was 1.86 with a 0.34 S.D. resulting in PERCIST thresholds of 2.54 and 2.87 when using 2 S.D. and 3 S.D, respectively. There were no noticeable differences between pre- and post- therapy background uptake. Examples of segmentations using the PERCIST thresholds are shown in Figure 4.11. The background mean and S.D. values in Table 4.11 were applied for all other segmentation methods using background uptake where applicable.

Dataset	VOI Size	Pre-Therapy Scans				Post-Therapy Scans			
		Mean	S.D.	PERCIST		Mean	S.D.	PERCIST	
				+2 S.D.	+3 S.D.			+2 S.D.	+3 S.D.
1	201ml	2.08	0.36	2.80	3.16	1.76	0.35	2.46	2.81
2	201ml	1.77	0.30	2.37	2.67	1.69	0.29	2.27	2.56
3	201ml	1.68	0.35	2.38	2.73	1.89	0.34	2.57	2.91
4	201ml	1.56	0.24	2.04	2.28	1.36	0.19	1.74	1.93
5	201ml	2.12	0.41	2.94	3.35	2.21	0.40	3.01	3.41
6	201ml	2.34	0.40	3.14	3.54	2.19	0.42	3.03	3.45
7	201ml	1.26	0.27	1.80	2.07	1.42	0.25	1.92	2.17
8	119ml	2.09	0.40	2.89	3.29	2.21	0.41	3.03	3.44
9	201ml	1.86	0.31	2.48	2.79	1.53	0.25	2.03	2.28
10	201ml	2.13	0.46	3.05	3.51	2.14	0.46	3.06	3.52
11	201ml	1.95	0.29	2.53	2.82	1.60	0.31	2.22	2.53
12	201ml	1.86	0.37	2.60	2.97	1.78	0.34	2.46	2.80
13	201ml	1.44	0.27	1.98	2.25	1.50	0.25	2.00	2.25
14	201ml	2.31	0.42	3.15	3.57	2.36	0.42	3.20	3.62
MEAN		**1.89**	**0.35**	**2.58**	**2.93**	**1.83**	**0.33**	**2.50**	**2.83**

Table 4.11: Values for Background Uptake for all 14 Datasets

The table shows the background region mean, S.D. and resulting PERCIST calculated segmentation thresholds for pre- and post- therapy images for each of the 14 datasets and the mean over all datasets. PERCIST thresholds are calculating using the background mean + 2 S.D. and 3 S.D. respectively.

4.4.8) Segmentation of Mesothelioma Patients using SUV_{max} and Background

Segmentation using 50% and 70% of the combination of SUV_{max} and background was used but produced high thresholds which only appeared to segment partial areas of disease (Figure 4.11). Lower percentages were used to try and solve this problem but caused major over-segmentation so this method was not used further in the segmentation of disease on the dataset.

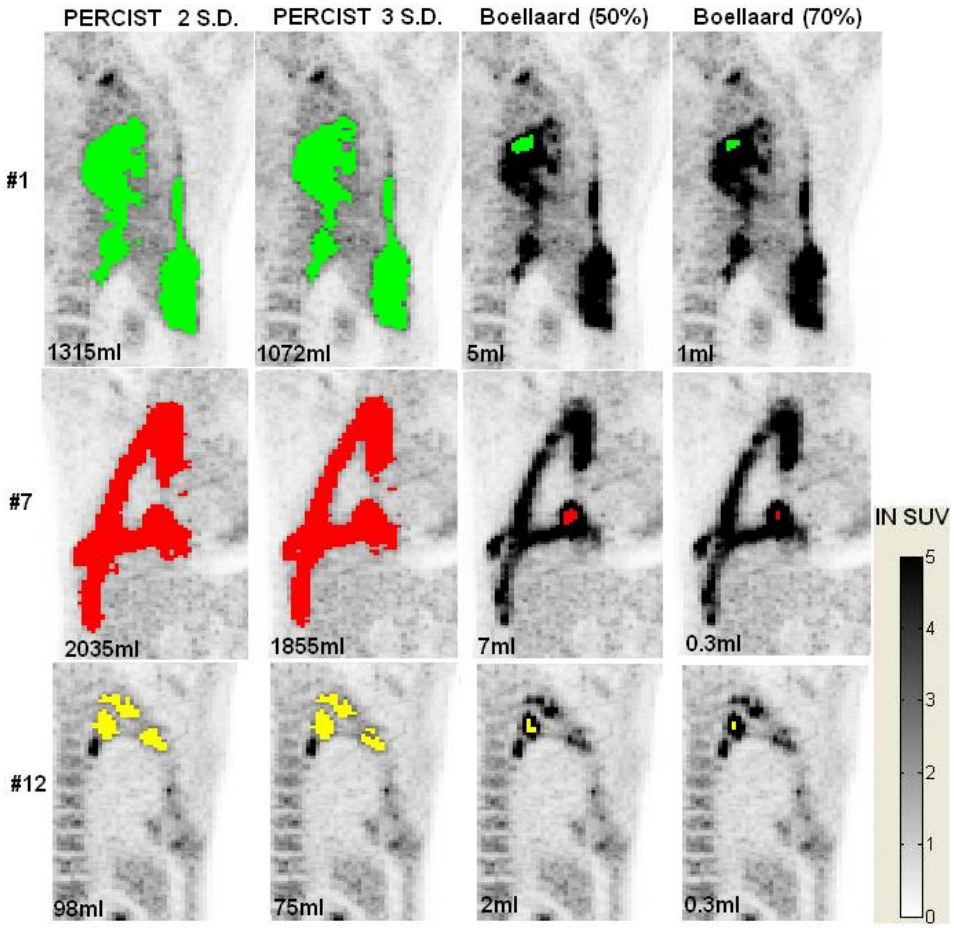


Figure 4.11: PERCIST and Boellaard Segmentation Methods

Segmentation of the primary disease region on pre-therapy scans for datasets 1, 7 and 12 using PERCIST segmentation threshold recommendations (Background Mean + 2 S.D./3 S.D) (Wahl *et al*., 2009), and 50% and 70% of SUV_{max} + Background Mean (Boellaard *et al*., 2004). TV is shown in the bottom left corner of each image. Images are of the coronal slice in which the SUV_{max} of the disease is present.

Another study using SUV_{max} and background, using a threshold based on a background corrected SUV_{max} calculation (Davis *et al*., 2006), was also applied to the mesothelioma data using different relative thresholds/percentages. Figure 4.12 shows examples of this method using different relative thresholds. A relative threshold of 30-50% appears too high to segment all disease, however, a lower relative threshold of 10% was used to segment the dataset with success.

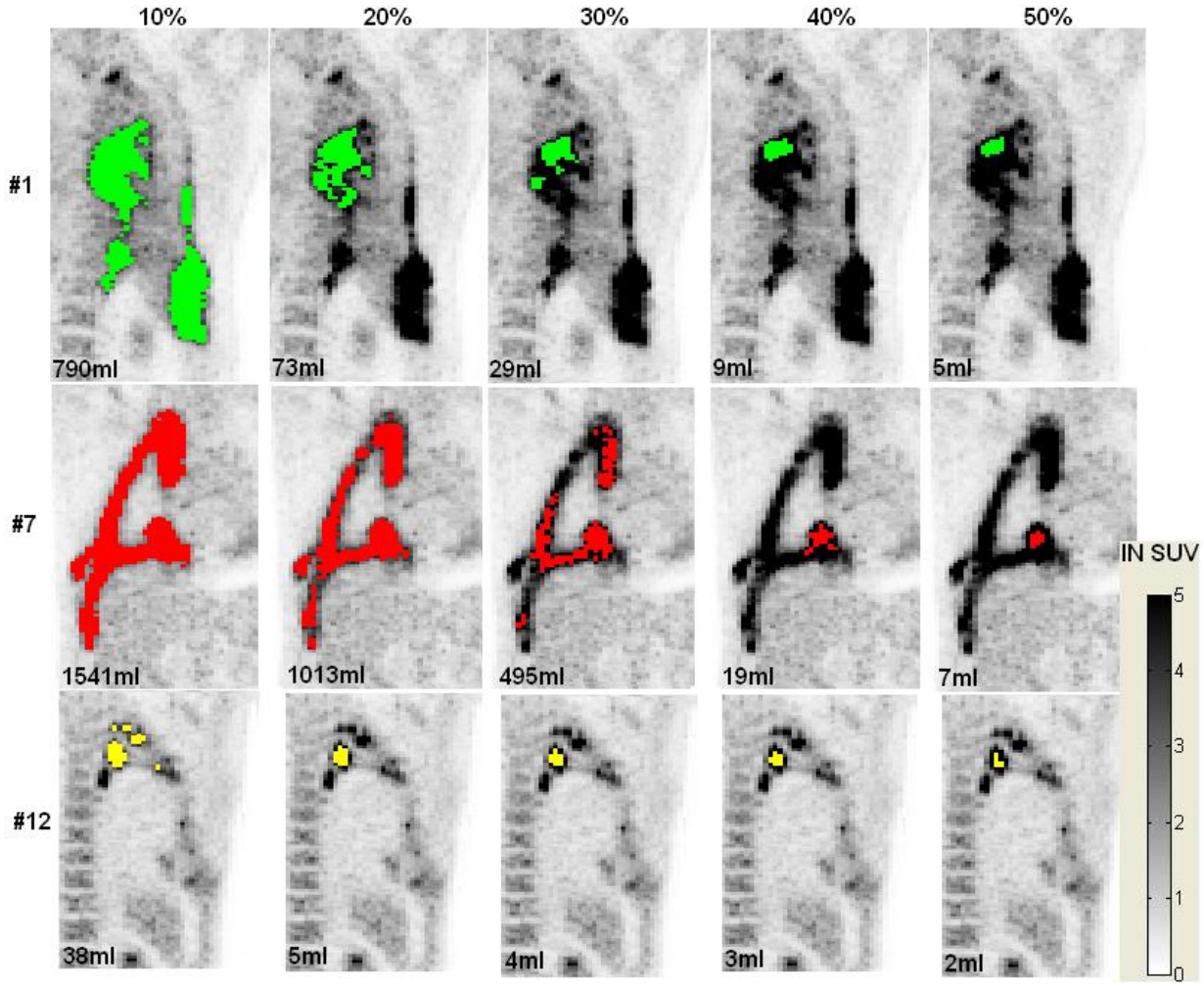


Figure 4.12: Background Corrected SUV_{max} Segmentation Method

Segmentation of the primary disease region on pre-therapy scans for datasets 1, 7 and 12 using background mean + % of SUV_{max} – background (Davis *et al*., 2006). 10%, 20%, 30% 40% and 50% were used as the relative threshold i.e. % in the equation (see 2.3.7 for more details, p103). TV is shown in the bottom left corner of each image. Images are of the coronal slice in which the SUV_{max} of the disease is present.

4.4.9) Segmentation of Mesothelioma Patients using SUV_{mean} and Background

Finally, the three segmentation methods using SUV_{mean} and background were tested on the mesothelioma patients, both threshold based calculations (Nestle *et al*., 2005; Nestle *et al*., 2007), and the adaptive GRAB threshold method (Boucek *et al*., 2008). Examples of segmentations using these methods are shown in Figure 4.13.

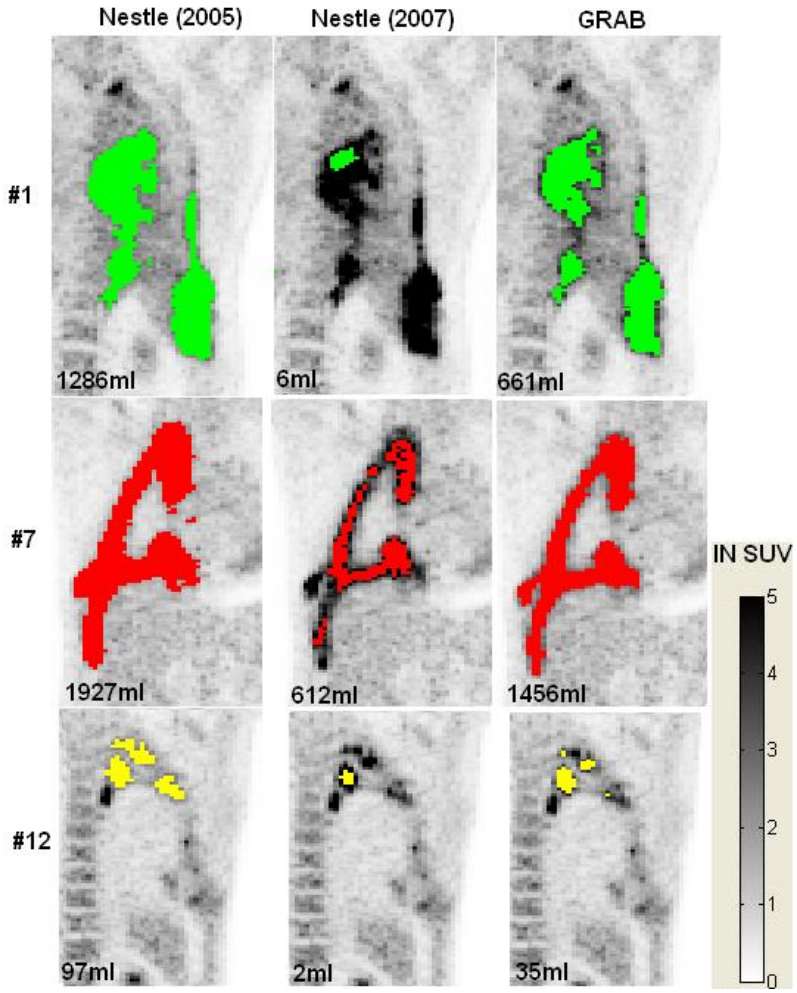


Figure 4.13: Segmentation Methods using Background and SUV_{mean}

Segmentation of the primary disease region on pre-therapy scans for datasets 1, 7 and 12 using threshold calculations of 0.15 * SUV_{mean} + Background (Nestle *et al*., 2005), (0.7 * SUV_{mean}) + (0.15 * Background) (Nestle *et al*., 2007), and the GRAB method using a threshold factor based on Background Mean + 3 S.D. (Boucek *et al*., 2008). TV is shown in the bottom left corner of each image. Images are of the coronal slice in which the SUV_{max} of the disease is present.

While two of the methods appeared to provide suitable segmentation, one threshold based calculation produced thresholds which appeared too high for the dataset (Nestle *et al*., 2007). Therefore, this method was not performed over the dataset of mesothelioma patients.

4.4.10) Comparison of Segmentation Methods on Mesothelioma Patients

With segmentation methods producing unrealistic thresholds removed, the nine remaining methods of segmentation were used on the dataset. They included: (i) the adaptive SUV_{mean} region growing method (Green *et al*., 2008), (ii) a SUV_{mean} threshold calculation method using three different combinations of fixed constants (Black *et al*., 2004), (iii) the recommended PERCIST threshold based on background uptake using both 2 S.D. and 3 S.D. (Wahl *et al*., 2009), (iv) a SUV_{mean} method using background uptake and a relative threshold (Davis *et al*., 2006), a calculation method using a percentage of SUV_{mean} and background uptake (Nestle *et al*., 2007) and the adaptive GRAB method using both SUV_{mean} and background uptake (Boucek *et al*., 2008).

While different fixed thresholds produced TV and TLG measurements which showed a low variability and high correlation, segmentation methods using calculations with varying parameters to decide upon a suitable threshold for each image individually, or in some cases for each tumour individually, can be expected to produce more variable results with lower correlations. A comparison of the number of segmented regions used throughout the 28 images, the number of times segmentation had to be restricted from background uptake and the mean values and changes between TV and TLG are displayed in Table 4.12. The results show that there are a low number of restricted areas across the segmentation methods, suggesting that the thresholds used are suitable as the segmentations do not require physiological uptake removing from disease, and the mean TV and TLG values are not drastically different to the fixed 2.5 SUV measure approved by a consultant physician. However, there are clear differences between the

segmentation methods which may have a positive or negative impact on TV and TLG measures in terms of accurately identifying response. The similarities with the physician approved 2.5 SUV threshold method would seem to suggest the inclusion of the nine segmentation methods is reasonable and they all have some merit in terms of segmenting disease in PET images.

Segmentation Method	Total VOIs	Restricted Areas	TV	TLG	Abs Change		Abs % Change	
					TV	TLG	TV	TLG
2.5 SUV	290	13	586	2593	154	840	82	89
Adaptive SUV_{mean}	152	3	187	1195	122	811	48	48
0.4 * SUV_{mean} + 1	270	8	507	2344	127	680	41	49
0.5 * SUV_{mean} + 0.6	261	13	533	2351	164	628	73	65
0.5 * SUV_{mean} + 1	358	3	311	1737	104	646	44	50
PERCIST 2 S.D.	312	7	636	2677	182	811	82	92
PERCIST 3 S.D.	369	5	523	2392	156	775	105	115
Davis *et al*. (2006)	254	6	459	2191	181	951	63	63
Nestle *et al*. (2005)	253	6	636	2684	175	801	45	53
GRAB	379	3	327	1799	121	778	125	143

Table 4.12: Difference between Non-Fixed Threshold Segmentation Algorithms

Table shows the total number of VOIs needed to segment disease, number of times restricted areas were used to stop segmentation, the mean TV (ml) and TLG (ml*SUV) for all 28 images and mean absolute change and absolute percentage change between pre- and post- therapy scans for all 14 datasets. All segmentation algorithms used region growing algorithms using 6-connected voxels. A fixed 2.5 SUV algorithm is included as a reference point.

As predicted, the correlation of the nine segmentation methods compared to a fixed 2.5 SUV threshold method was not as high as other fixed threshold methods and there was often very little correlation in the change and percentage change between pre- and post- therapy scans (pcc ranging from 0.109 to 0.943 for TV and 0.219 to 0.974 for TLG, Table 4.13). The percentage change between pre- and post- therapy scans can be used as a response measure and the lack of correlation between segmentation methods suggest that the method chosen could have a significant influence on its predictability. Of the segmentation methods, it is interesting to note

that the PERCIST recommended threshold calculations correlate highly with SUV (pcc = 0.943, p = 0.0001 and pcc = 0.867, p = 0.0001 for percentage change in TV). This is not surprising given the mean thresholds were ~2.5 SUV over the dataset and that the method does use a fixed threshold across an image.

Segmentation Method	Over all 28 Images		Change		% Change	
	TV	TLG	TV	TLG	TV	TLG
Fixed 2.5 SUV	1.000	1.000	1.000	1.000	1.000	1.000
Adaptive SUV_{mean}	0.877	0.927	0.846	0.954	0.282	0.451
0.4 * SUV_{mean} + 1	0.969	0.993	0.675	0.978	0.686	0.768
0.5 * SUV_{mean} + 0.6	0.864	0.974	0.070	0.895	0.160	0.219
0.5 * SUV_{mean} + 1	0.974	0.990	0.897	0.983	0.109	0.259
PERCIST 2 S.D.	0.946	0.984	0.677	0.962	0.943	0.974
PERCIST 3 S.D.	0.942	0.981	0.823	0.977	0.867	0.903
Davis *et al*. (2006)	0.920	0.965	0.879	0.955	0.188	0.369
Nestle *et al*. (2005)	0.950	0.985	0.416	0.920	0.475	0.654
GRAB	0.927	0.970	0.892	0.982	0.781	0.836

Table 4.13: Correlation between Non-Fixed Threshold Segmentation Methods

Pearson correlation coefficients (pccs) are shown for the correlation of various non-fixed threshold segmentation methods with a 2.5 SUV fixed threshold for TV and TLG (correlation with SUV_{max} was 1.000 in all cases). Pccs were calculated for TV and TLG over the 28 pre- and post- therapy images and for the change and percentage change between the 14 pre- and post- therapy scans. For $p < 0.01$, $pcc > 0.479$ over the 28 images and for $p < 0.01$, $pcc > 0.662$ for changes and percentage change.

To highlight the potential difference the segmentation method can make, Tables 4.14 and 4.15 show the correlation between each of the nine segmentation methods for the percentage change between pre- and post- therapy for TV and TLG between, respectively. The correlation varies from strong correlations to no correlation amongst the nine segmentation methods for both TV and TLG so while using some segmentation methods may not result in significant changes in terms for the TV or TLG measures, other methods may influence this considerably. The effect on response is analysed in the 4.5.5 and 4.5.6. However, it can be conclude that the segmentation

method has an influence on TV and therefore, when assessing the success of TV and TLG at predicting response, the segmentation method used to obtain these values must be taken into consideration.

Segmentation Methods	(A) Adapt Mean	(B) 0.4*M +1	(C) 0.5*M +0.6	(D) 0.5*M +1	(E) PERC 2 S.D.	(F) PERC 3 S.D.	(G) Davis 2006	(H) Nestle 2005	(I) GRAB
(A) Adapt SUV_{mean}	1.000	0.169	0.391	0.584	0.200	0.239	0.685	0.136	0.144
(B) 0.4 * Mean + 1	0.169	1.000	0.374	0.491	0.696	0.668	0.395	.698	0.691
(C) 0.5 * Mean + 0.6	0.391	0.374	1.000	0.355	-0.057	-0.092	0.273	-0.179	-0.064
(D) 0.5 * Mean + 1	0.584	0.491	0.355	1.000	0.216	0.323	0.719	0.636	0.375
(E) PERCIST 2 S.D.	0.200	0.696	-0.057	0.216	1.000	0.965	0.227	0.703	0.912
(F) PERCIST 3 S.D.	0.239	0.668	-0.092	0.323	0.965	1.000	0.282	0.788	0.964
(G) Davis (2006)	0.685	0.395	0.273	0.719	0.227	0.282	1.000	0.446	0.280
(H) Nestle (2005)	0.136	0.698	-0.179	0.636	0.703	0.788	0.446	1.000	0.817
(I) GRAB	0.144	0.691	-0.064	0.375	0.912	0.964	0.280	0.817	1.000

Table 4.14: Correlation of Percentage Change in Tumour Volume between Segmentations

Pearson correlation coefficients (pccs) are shown for the correlation of percentage change in TV between all non-fixed threshold segmentation methods in all 14 pre- and post- therapy scans. For $p < 0.01$, pcc > 0.479 over the 28 images and for $p < 0.01$, pcc > 0.662 for changes and percentage change.

Segmentation Methods	(A) Adapt Mean	(B) 0.4*M +1	(C) 0.5*M +0.6	(D) 0.5*M +1	(E) PERC 2 S.D.	(F) PERC 3 S.D.	(G) Davis 2006	(H) Nestle 2005	(I) GRAB
(A) Adapt SUV_{mean}	1.000	0.489	0.349	0.648	0.396	0.410	0.775	0.432	0.303
(B) 0.4 * Mean + 1	0.489	1.000	0.446	0.669	0.782	0.786	0.661	0.838	0.802
(C) 0.5 * Mean + 0.6	0.349	0.446	1.000	0.365	0.072	0.015	0.368	0.013	0.035
(D) 0.5 * Mean + 1	0.648	0.669	0.365	1.000	0.322	0.437	0.857	0.742	0.461
(E) PERCIST 2 S.D.	0.396	0.782	0.072	0.322	1.000	0.968	0.397	0.776	0.922
(F) PERCIST 3 S.D.	0.410	0.786	0.015	0.437	0.968	1.000	0.468	0.873	0.976
(G) Davis (2006)	0.775	0.661	0.368	0.857	0.397	0.468	1.000	0.671	0.454
(H) Nestle (2005)	0.432	0.838	0.013	0.742	0.776	0.873	0.671	1.000	0.893
(I) GRAB	0.303	0.802	0.035	0.461	0.922	0.976	0.454	0.893	1.000

Table 4.15: Correlation of Percentage Change in Total Lesion Glycolysis between Segmentations

Pearson correlation coefficients (pccs) are shown for the correlation of percentage change in TLG between all non-fixed threshold segmentation methods in all 14 pre- and post- therapy scans. For $p < 0.01$, pcc > 0.479 over the 28 images and for $p < 0.01$, pcc > 0.662 for changes and percentage change.

4.5) Predicting Survival using Response Measures

4.5.1) Measures of Survival

For the 14 patients with MPM, PFS and OS were chosen as parameters with which to measure response, with longer PFS and OS times indicating patients with good response to therapy. Any response parameter which can accurately predict good or bad PFS and/or OS would be of particular interest. OS and PFS of the 14 patients are listed in Table 4.16. All patients were found to have disease progression and died before data was collected with the exception of patient 14 who did not die or have disease progression confirmed 800 days after the start of treatment. 800 days was the 2nd longest OS time and longest PFS time so while there is no definite endpoint for this patient it was not thought that it would significantly alter results as the patient had already had a long survival time, in comparison to others, at the last time of follow-up.

4.5.2) Statistical Methods for Assessing Response

Three different methods of statistical analysis were used to assess the success both qualitative and quantitative parameters. Firstly, Pearson correlation coefficients (pcc) were used to see how well the response measures correlated with OS and PFS, with well correlated measures potentially being accurate response measures because they are heavily related to the survival time of patient. However, just because a parameter correlates well with survival, it does not necessairily mean that it will be useful in a clinical setting without a way of diving patients into groups of responders and non-responders. Secondly, receiver operator characteristics (ROC) curves were used in an attempt to detect the sensitivity and specificity of a measure at predicting the status of a patient. ROC curves attempt to define the accuracy of a response measure by defining the ability of a response method for having true positives or negatives. Using ROC analysis allows the most appropriate threshold to be obtained for dividing patients into groups of responders and non-responders with the best discrimination between groups. The area under the curve (AUC)

helps define whether any threshold of a particular measure can discriminate groups of patients better than random guessing.

If ROC analysis showed a parameter could divide patients into responders and non-responders successfully, Kaplan-Meier survival curves were plotted to analyse the significance of the two groups OS and PFS. However, with a sample size of only 14 patients, even very strong correlations or ROC curves with high AUC have low predictive power and statistical significance. For example, for the pcc to be statistically significant ($p < 0.05$) for 14 patients, correlation would have to be >0.533. Conclusions made must take into account the small sample size of the dataset. However, results should at least show what measures have the potential to be effective even if more studies with more patients are needed to determine this further. The potential clinical benefit of a response measure which could accurately distinguish between responders and non-responders is that non-responding patients could be taken off a therapy regime which is not resulting in a response and given an alternative treatment.

For ROC analysis, the OS and PFS data must be in a binary format so the predicted response parameter can predict whether a patient has, for example, survived or not. Therefore, a threshold is used for OS and PFS to determine whether a patient has survived for a long or short period of time. Thresholds for studies looking at the survival of patients in medical scenarios can be variable, however, in MPM patients OS or PFS is relatively low so an OS or PFS of over six months or a year is seen as a positive outcome. With 14 patients, it is also important to make sure that patients are relatively uniformly separated into two groups. Therefore, for ROC analysis a threshold of six months for OS and PFS was used to differentiate responders and non-responders with nine patients alive and five patients deceased for OS at six months and eight patients alive and six patients deceased or found to have disease progression for PFS. As the measures are

similar, it is not expected that results will differ significantly. Statistical analysis was done using statistical software package SPSS Statistics Version 20 (IBM, Armonk, NY, USA).

4.5.3) Predicting Survival in Patients with Mesothelioma using Visual Analysis

The visual analysis responses using both custom-made PET criteria (Table 4.2, p185) and modified RECIST (Table 4.3, p185) are listed for all 14 patients in Table 4.16 along with OS and PFS. By viewing this table, it can be seen that visual analysis does not predict either OS or PFS. There are only three patients classified as responders using visual analysis on PET and all three have short survival times as opposed to longer survival times that one would expect for responding patients. Visual analysis on CT only has one patient as partially responding and while they have a reasonably long survival time, other patients with longer survival are classified as having stable disease. There is some correlation between CT visual analysis and OS and PFS (pcc = 0.405, p = 0.15 and pcc = 0.209, p = 0.48, respectively with each category of response given as a numerical index i.e. 1 = PR, 2 = SD, 3 = PD), however, it is not statistically significant. Similarly, PET visual analysis did not have any statistically significant correlation. ROC analysis and Kaplan-Meier survival curves were used to investigate both visual analysis methods but these results only confirmed that the two measures did not predict response.

Figure 4.14 shows the Kaplan-Meier survival curves for OS and PFS for all 14 patients. The curves show that there are low levels of survival using both measures. The one patient who did not die or progress in the study is marked with a + on the curve, whereas all other patients were deceased at the time of follow-up. A good response measure should produce two curves for two groups of patients, one in which cumulative survival decreases rapidly for non-responders and one in which cumulative survival remains steady for good responders.

Dataset	Survival Parameters		Visual Response Assessment	
	OS (days)	PFS (days)	CT Response (RECIST)	PET Response
1	164	96	Stable Disease	Response
2	155	48	Progressive Disease	Response
3	577	229	Partial Response	Progressive Metabolic Disease
4	515	247	Stable Disease	Stable Metabolic Disease
5	1070	196	Stable Disease	Progressive Metabolic Disease
6	250	217	Stable Disease	Progressive Metabolic Disease
7	158	97	Stable Disease	Response
8	477	244	Stable Disease	Progressive Metabolic Disease
9	132	68	Progressive Disease	Progressive Metabolic Disease
10	750	286	Stable Disease	Stable Metabolic Disease
11	418	221	Stable Disease	Progressive Metabolic Disease
12	545	114	Stable Disease	Stable Metabolic Disease
13	152	13	Stable Disease	Progressive Metabolic Disease
14	800	800	Stable Disease	Progressive Metabolic Disease
Mean	440	205	N/A	N/A

Table 4.16: Survival and Visual Response for Mesothelioma Patients

Table shows OS and PFS, in days, for all 14 mesothelioma patients in the dataset. Patient 14 was the only patient still alive when the data was collected, 800 days after the start of patient 14's therapy start date, and the disease had not been found to progress. PET and CT response were judged based on custom-made criteria (Table 4.2, p185) and modified RECIST (Table 4.3, p185, (Byrne and Nowak, 2004)), respectively.

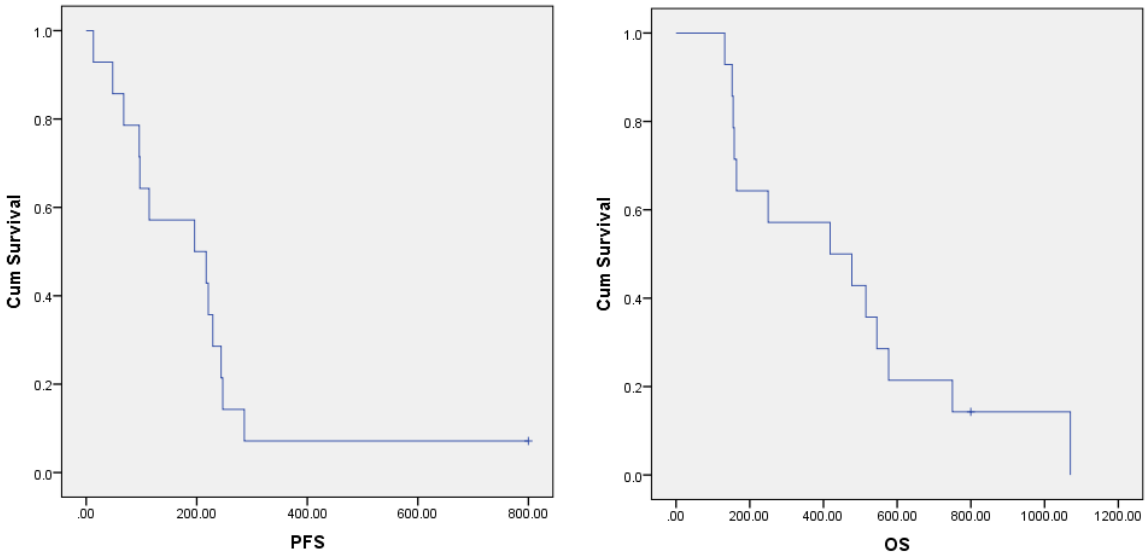


Figure 4.14: Kaplan-Meier Survival Curves for all Mesothelioma Patients

Survival curves for (A) PFS and (B) OS for all 14 mesothelioma patients. OS and PFS are in days.

4.5.4) Predicting Survival in Patients with Mesothelioma using SUV_{max} and SUV_{peak}

In investigating response based on SUV_{max} and SUV_{peak} measurements, a number of factors must be taken into consideration. All parameters are taken before and after therapy so response can be based on measurements before therapy, after therapy or the change between the two values as either an absolute or percentage change. It is generally considered that the change from pre-therapy to post-therapy, usually as a percentage, is, theoretically, the best predictor of response considering it shows the impact treatment has had on disease. However, this is not necessarily the case. The change from pre- to post- therapy is noted as a change (or percentage change) with negative values indicating a reduction from pre- to post- therapy, therefore, suggesting a good response, while positive values indicate an increase from pre- to post- therapy suggesting a poor response. If TV has percentage change of -5% from pre- to post- therapy scans then it has reduced in size and should indicate a response in the patient compared to a change of +5%.

Maximum uptake values were obtained with different methods of calculating the total volume (SUV, SUL and SUV_{BSA}) and peak values were obtained using both PETTRA and HERMES software. As expected there is excellent correlation between SUV_{max}, SUL_{max} and SUV_{BSA}, both in terms of pre- and post- therapy values and the change and percentage change between them. The lowest correlation is still high, between the change in pre- and post- therapy values, between SUV_{max} and SUV_{BSA} (pcc = 0.968, $p < 0.0001$). SUV_{peak} values taken by HERMES and PETTRA are also very highly correlated with the lowest correlation being the percentage change between pre- and post- therapy parameters (pcc = 0.970, $p < 0.0001$). The correlation of all the parameters with each other for pre- and post- therapy values and the change and percentage change between them is shown in Table 4.15. The table shows very high correlation between values, suggesting that the difference in predicting response between them will be negligible. However, there does seem a weaker correlation between the SUV_{max} and SUV_{peak} parameters.

(A) Pre- and Post-Therapy Values	SUV_{max}	SUL_{max}	$SUV_{BSA\ max}$	PETTRA SUV_{peak}	HERMES SUV_{peak}
SUV_{max}	1.000	0.993	0.978	0.955	0.954
SUL_{max}	0.993	1.000	0.979	0.940	0.942
SUV_{BSAmax}	0.978	0.979	1.000	0.955	0.958
PETTRA SUV_{peak}	0.955	0.940	0.955	1.000	0.996
HERMES SUV_{peak}	0.954	0.942	0.958	0.996	1.000

(B) Change between Pre- and Post- Therapy	SUV_{max}	SUL_{max}	$SUV_{BSA\ max}$	PETTRA SUV_{peak}	HERMES SUV_{peak}
SUV_{max}	1.000	0.994	0.968	0.742	0.761
SUL_{max}	0.994	1.000	0.984	0.698	0.728
SUV_{BSAmax}	0.968	0.984	1.000	0.638	0.669
PETTRA SUV_{peak}	0.742	0.698	0.638	1.000	0.975
HERMES SUV_{peak}	0.761	0.728	0.669	0.975	1.000

(C) % Change between Pre- and Post- Therapy	SUV_{max}	SUL_{max}	$SUV_{BSA\ max}$	PETTRA SUV_{peak}	HERMES SUV_{peak}
SUV_{max}	1.000	0.996	0.990	0.849	0.873
SUL_{max}	0.996	1.000	0.995	0.842	0.876
SUV_{BSAmax}	0.990	0.995	1.000	0.832	0.858
PETTRA SUV_{peak}	0.849	0.842	0.832	1.000	0.970
HERMES SUV_{peak}	0.873	0.876	0.858	0.970	1.000

Table 4.17: Correlation between Max and Peak Response Measures

Pearson correlation coefficients (pccs) are shown for each SUV_{max} and SUV_{peak} measure. Pccs were calculated for both pre- and post-therapy values for each of the response parameters (Table A) and the change (Table B) and percentage change (Table C) between them over the 14 pre- and post- therapy scans. For $p < 0.01$, pcc > 0.662 and $p < 0.05$, pcc > 0.533. The only difference between the SUV_{max}, SUL_{max} and SUV_{BSAmax} was in the units used and not in the formulation of the maximum value in TV.

Both pre- and post- therapy values for all five parameters correlated well with OS and PFS with pcc ranging from -0.467 to -0.702 ($p = 0.1$ to 0.005) while the change and percentage change between them had much weaker correlation with OS with pcc ranging from 0.191 to 0.480 ($p = 0.5$ to 0.08) and no correlation with PFS with pcc ranging from 0.091 to 0.199 (Table 4.18). As

would be expected, a high pre- or post- therapy SUV_{max} correlates with a low OS or PFS. SUL and SUV_{BSA} show mildly better correlations with OS and PFS than for SUV_{max} but the three are very similar. Pre- and post- therapy SUV_{peak} values, both from PETTRA and HERMES, have the strongest correlation with OS and PFS, the difference between the two is minimal. The change between pre- and post- therapy values for all five parameters produces a correlation where a negative change from pre- to post- therapy correlates with a low OS or PFS when a negative change is supposed to be an indicator of response. This is a surprising result as one would expect that, as with the pre- and post- therapy values, they should correlate negatively with OS and PFS.

Response Parameter	OS				PFS			
	Pre	Post	Change	% Chg	Pre	Post	Change	% Chg
SUV_{max}	-0.467	-0.507	0.278	0.422	-0.579	-0.650	0.199	0.139
SUL_{max}	-0.507	-0.528	0.237	0.404	-0.575	-0.652	0.147	0.104
SUV_{BSAmax}	-0.526	-0.548	0.191	0.397	-0.591	-0.653	0.122	0.091
PETTRA SUV_{peak}	-0.620	-0.527	0.468	0.414	-0.606	-0.702	0.100	-0.091
HERMES SUV_{peak}	-0.606	-0.540	0.466	0.480	-0.587	-0.699	0.108	-0.091

Table 4.18: Correlation between Max and Peak Response Parameters and Survival

Pearson correlation coefficients (pccs) are shown for the correlation of each SUV_{max} and SUV_{peak} response measure with OS and PFS for pre-therapy values, post-therapy values, and change and percentage change (chg) from pre- to post- therapy values. For $p < 0.01$, $pcc > 0.662$ and $p < 0.05$, $pcc > 0.533$.

This is confirmed further when using ROC curves to investigate the parameters predicting PFS and OS at 6 months. Pre- and post- therapy SUV_{max} produce excellent ROC curves which predict response accurately, however, the change and percentage change do the inverse of this (Figure 4.15). It is difficult to determine an exact reason for the inverse correlation of change and percentage change, however, studying the data, it appears that higher SUV_{max} values are more likely to have a negative change between pre- and post- therapy scans than lower SUV_{max} values. For example, for dataset 1 there is a change of -1.6 SUV, however, the original pre-therapy SUV of 18.5 is high so it is unsurprising the patient had poor OS of 164 days. In comparison, a low

SUV_{max} in dataset 5 changes from 6 SUV to 9.2 SUV, however, with a much lower SUV_{max} for both values it is hardly surprising that the patient went on the have an OS of 1070 days. While the positive correlation with OS and PFS for the change in SUV_{max} is a surprising result, it is explainable.

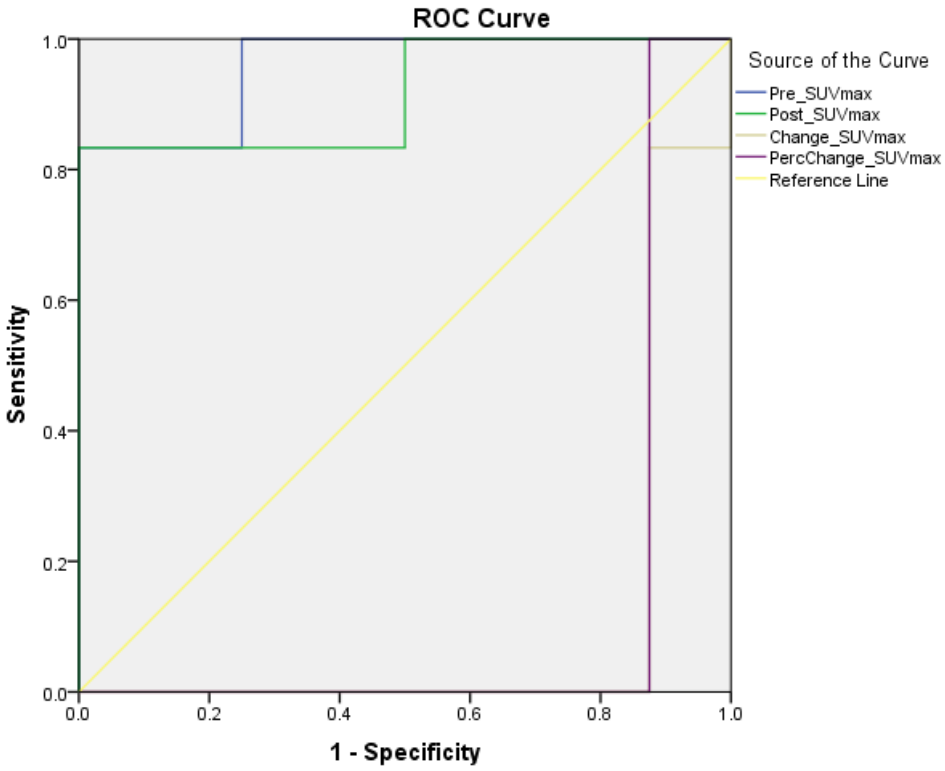


Figure 4.15: ROC Curve for SUV_{max} Parameters for PFS at 6 months

ROC curves are plotted for pre-therapy SUV_{max} , post-therapy SUV_{max}, change in SUV_{max} and percentage change in SUV_{max}. The curves show that pre- and post- therapy SUV_{max} provide strong ROC curves while change and percentage change in SUV_{max} have unwanted results. ROC curves for all five maximum and peak response measures are very similar. Pre-therapy SUV_{max} has an AUC of 0.844 (Standard Error (SE) = 0.113, 95% Confidence Intervals (CI) of 0.622 – 1). Post-therapy SUV_{max} has an AUC of 0.800 (SE = 0.131, 95% CI = 0.544 – 1).

There is very little difference between using SUV_{max}, SUL_{max}, SUV_{BSA}, PETTRA SUV_{peak} or HERMES SUV_{peak} for predicting response. As shown in Table 4.18, their correlations are similar as are their ROC curves, highlighted in Table 4.19, which shows the AUC for each of parameters ROC curves, all proving to have very similar values.

Response Parameter	6 Month OS				6 Month PFS			
	Pre	Post	Change	% Chg	Pre	Post	Change	% Chg
SUV_{max}	0.844	0.800	0.222	0.222	0.958	0.917	0.104	0.125
SUL_{max}	0.867	0.822	0.222	0.222	0.979	0.937	0.104	0.125
SUV_{BSAmax}	0.867	0.800	0.267	0.222	0.979	0.917	0.146	0.125
PETTRA SUV_{peak}	0.844	0.778	0.178	0.200	0.958	0.896	0.104	0.146
HERMES SUV_{peak}	0.844	0.778	0.178	0.222	0.958	0.896	0.104	0.188

Table 4.19: AUC for Max and Peak Response Parameters Predicting Survival

Area under the curve (AUC) is shown for the ROC curves for all five response parameters for predicting six month OS and PFS for pre- and post- therapy values and change and percentage change between them.

The results show there is no significant difference between the max and peak parameters when it comes to predicting OS and PFS. Also, pre- or post- therapy values alone are more valuable than investigating the change and percentage change between them in terms of predicting response for this dataset. A pre-therapy SUV_{max} of 13 is the best threshold for ROC analysis, however, as SUV is heavily dependent on the scanner and scanning protocol using this value for other studies, particularly given the small number of patients in this study, is not advisable. The Kaplan-Meier survival curves for two groups of patients split by a pre-therapy SUV_{max} of 13 are shown in Figure 4.16. While this is statistically significant ($p < 0.001$), it should be considered that with only 14 patients, there is more of a possibility that this has happened by chance than because the measure is a robust measure of distinguishing patients with significantly different survival times.

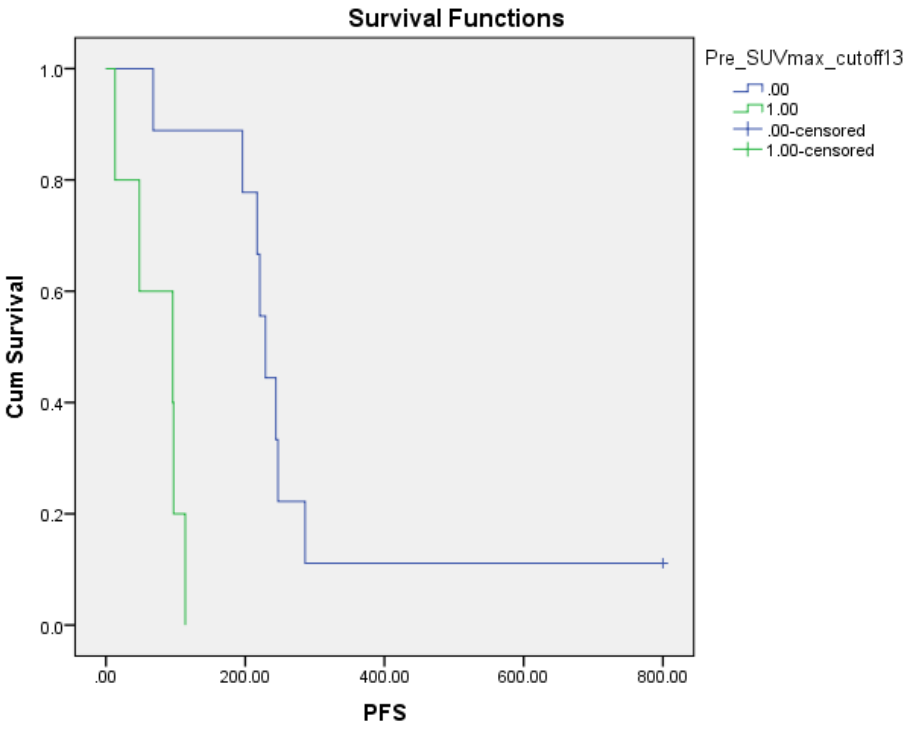


Figure 4.16: Kaplan-Meier Survival Curve for Pre-Therapy SUV_{max} and PFS

Survival curves are plotted for two groups of patients, those with pre- therapy SUV_{max} > 13 SUV (Mean PFS = 74, SE = 19, 95% CI = 37-110) and those < 13 SUV (mean PFS = 279, SE = 64, 95% CI = 153-405) Comparing the survival distributions using the log-rank method shows a significance of $p < 0.001$.

4.5.5) Predicting Survival in Patients with Mesothelioma using Tumour Volume and Total Lesion Glycolysis using Fixed Threshold Segmentation

The segmentation method is an important factor in assessing response using TV and TLG. A variety of fixed and non-fixed threshold methods have been used to obtain TV and TLG in this study. The correlation between fixed threshold methods was found to be high while the correlation between non-fixed threshold methods is more variable. Therefore, different fixed threshold methods are expected to have similar results in terms of predicting OS and PFS while non-fixed methods are expected to have more of a difference. Segmentation using 6-connected and 26-connected voxels was found to correlate so well that the differences in terms of identifying response has not been investigated as it is thought to be inconsequential. The

correlation of TV and TLG with OS and PFS for pre- and post- therapy values and the change and percentage change between them are shown in Table 4.20, for TV, and Table 4.21, for TLG.

Segmentation Methods for TV	OS				PFS			
	Pre	Post	Change	% Chg	Pre	Post	Change	% Chg
SUV 2.2	-0.641	-0.622	0.103	0.264	-0.430	-0.458	-0.095	0.019
SUV 2.5	-0.642	-0.638	0.064	0.195	-0.419	-0.447	-0.059	-0.038
SUV 3	-0.635	-0.633	0.154	0.319	-0.405	-0.429	0.049	-0.148
SUL 2.2	-0.643	-0.649	0.012	0.304	-0.411	-0.441	-0.076	-0.064
SUL 2.5	-0.634	-0.644	0.111	0.327	-0.402	-0.428	0.026	-0.019
SUL 3	-0.617	-0.625	0.173	0.455	-0.386	-0.402	0.093	0.002
SUV_{BSA} 0.5	-0.720	-0.690	-0.058	0.449	-0.480	-0.525	-0.288	0.083
SUV_{BSA} 0.6	-0.668	-0.678	-0.228	0.189	-0.439	-0.476	-0.289	0.020
SUV_{BSA} 0.7	-0.651	-0.673	-0.096	0.297	-0.423	-0.459	-0.135	-0.064

Table 4.20: Correlation between TV with Fixed Segmentation Thresholds and Survival

Pearson correlation coefficients (pccs) are shown for the correlation of TV, for each fixed threshold segmentation method, with OS and PFS for pre-therapy values, post-therapy values, and the change and percentage change (chg) from pre- to post- therapy values. For $p < 0.01$, pcc > 0.662 and $p < 0.05$, pcc > 0.533.

Segmentation Methods for TLG	OS				PFS			
	Pre	Post	Change	% Chg	Pre	Post	Change	% Chg
SUV 2.2	-0.639	-0.626	0.203	0.296	-0.419	-0.439	0.067	0.009
SUV 2.5	-0.637	-0.627	0.178	0.238	-0.411	-0.424	0.075	-0.034
SUV 3	-0.631	-0.613	0.202	0.362	-0.402	-0.404	0.107	-0.125
SUL 2.2	-0.631	-0.634	0.130	0.336	-0.402	-0.419	0.052	-0.054
SUL 2.5	-0.624	-0.623	0.162	0.346	-0.396	-0.405	0.087	-0.015
SUL 3	-0.613	-0.600	0.191	0.464	-0.387	-0.380	0.118	0.005
SUV_{BSA} 0.5	-0.689	-0.683	0.079	0.489	-0.455	-0.492	-0.079	0.079
SUV_{BSA} 0.6	-0.651	-0.666	0.031	0.210	-0425	-0.455	-0.030	0.016
SUV_{BSA} 0.7	-0.638	-0.654	0.062	0.319	-0.414	-0.438	0.013	-0.056

Table 4.21: Correlation between TLG with Fixed Segmentation Thresholds and Survival

Pearson correlation coefficients (pccs) are shown for the correlation of TLG, for each fixed threshold segmentation method, with OS and PFS for pre-therapy values, post-therapy values, and the change and percentage change (chg) from pre- to post- therapy values. For $p < 0.01$, pcc > 0.628 and $p < 0.05$, pcc > 0.497.

The results in both Table 4.20 and 4.21 indicate that TV and TLG are very similar when investigating response, as are the different segmentation methods for obtaining them. Correlations are very similar for each segmentation method and between the TV and TLG parameters themselves. Once again, results show that the pre- and post- therapy values both correlate with OS significantly. However, the change and percentage change between them suffer the same issues with SUV_{max} and SUV_{peak} values and often show very little correlation or no correlation at all. This is also true for PFS and while pre- and post- therapy TV and TLG show a significant correlation with OS, there is insignificant correlation with PFS. When assessing ROC curves a similar pattern is seen, both pre- and post- therapy TV appear to predict both six month OS and PFS with an AUC between 0.750 – 0.911, while the change between TV does not appear to predict OS and PFS in any meaningful way (Table 4.22).

Segmentation Method for TV	OS				PFS			
	Pre	Post	Change	% Chg	Pre	Post	Change	% Chg
SUV 2.2	0.867	0.844	0.422	0.444	0.875	0.750	0.313	0.375
SUV 2.5	0.867	0.889	0.511	0.489	0.875	0.812	0.396	0.438
SUV 3	0.889	0.889	0.444	0.467	0.875	0.833	0.354	0.396
SUL 2.2	0.889	0.889	0.578	0.489	0.875	0.833	0.500	0.438
SUL 2.5	0.867	0.867	0.533	0.444	0.875	0.833	0.417	0.375
SUL 3	0.867	0.867	0.444	0.400	0.875	0.833	0.354	0.354
SUV_{BSA} 0.5	0.933	0.911	0.600	0.356	0.896	0.833	0.521	0.271
SUV_{BSA} 0.6	0.911	0.911	0.600	0.400	0.875	0.833	0.500	0.333
SUV_{BSA} 0.7	0.911	0.889	0.622	0.422	0.875	0.854	0.500	0.354

Table 4.22: AUC for TV for Fixed Segmentation Methods for Predicting Survival

Area under the curve (AUC) is shown for the ROC curves for TV using nine different fixed segmentation thresholds for predicting six month OS and PFS for pre- and post- therapy values and the change and percentage change between them.

The results for TLG provide very similar results to TV with pre- and post- therapy TLG predicting both six month OS and PFS with an AUC ranging from 0.812 – 0.911 (Table 4.23). Once again, the change and percentage change between pre- and post- therapy have no real

predictive power and as with SUV_{max}, occasionally a negative change in TV and TLG actually correlates with a poor prognosis. This is presumably for the same reason as with SUV_{max} where by greater changes in TV and TLG are seen in higher TVs and TLGs. Of all the fixed threshold segmentation methods, SUV_{BSA} 0.5 has the best AUC on ROC analysis and correlations. This is worth considering, however, all the fixed segmentation methods produce similar results and further assessment on larger datasets would be needed to see which produces TVs most beneficial for predicting survival.

Segmentation Method for TLG	OS				PFS			
	Pre	Post	Change	% Chg	Pre	Post	Change	% Chg
SUV 2.2	0.867	0.889	0.311	0.400	0.875	0.812	0.229	0.333
SUV 2.5	0.867	0.867	0.400	0.444	0.875	0.833	0.313	0.396
SUV 3	0.867	0.867	0.444	0.378	0.875	0.833	0.354	0.333
SUL 2.2	0.867	0.867	0.422	0.444	0.875	0.833	0.333	0.375
SUL 2.5	0.867	0.867	0.444	0.356	0.875	0.833	0.354	0.313
SUL 3	0.867	0.867	0.444	0.400	0.875	0.833	0.354	0.354
SUV_{BSA} 0.5	0.911	0.889	0.467	0.289	0.875	0.833	0.375	0.208
SUV_{BSA} 0.6	0.911	0.889	0.578	0.400	0.875	0.854	0.458	0.333
SUV_{BSA} 0.7	0.889	0.867	0.467	0.400	0.875	0.833	0.375	0.313

Table 4.23: AUC for TLG for Fixed Segmentation Methods for Predicting Survival

Area under the curve (AUC) is shown for the ROC curves for TLG using nine different fixed segmentation thresholds for predicting six month OS and PFS for pre- and post- therapy values and the change and percentage change between them.

To show the similarities between the segmentation methods, pre-therapy TLG ROC curves for six month OS are shown for segmentations using thresholds of SUV 2.2, SUV 3, SUL 2.2, SUL 3, SUV_{BSA} 0.5 and SUV_{BSA} 0.7 (Figure 4.17). The ROC curves for six month PFS are even more similar, in fact, they are identical. A Kaplan-Meier survival curve for PFS grouping the two patients into groups based on a cut-off of ~1300ml TV based on a SUV 2.5 threshold shows the potential pre-therapy TV has in predicting response (Figure 4.18). While there is clearly some

divide between these patients, only three have a pre-therapy value below the cut-off and it is not statistically significant ($p < 0.021$).

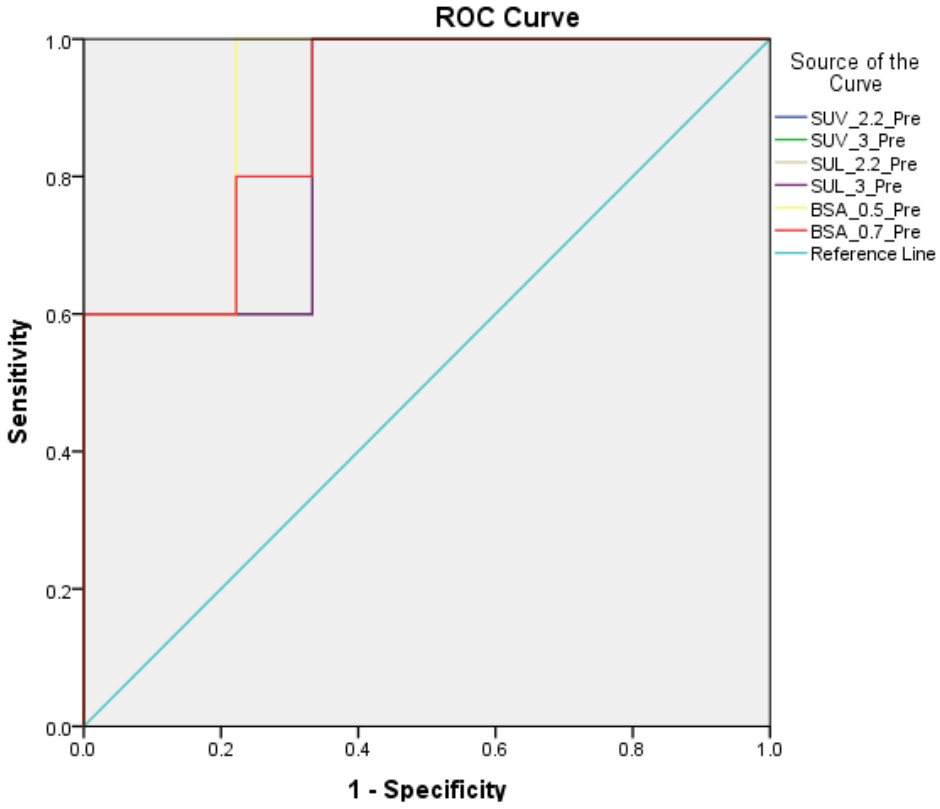


Figure 4.17: ROC Curve for Pre-Therapy TLG for PFS at 6 months

ROC curves are plotted for pre-therapy TLG for six fixed segmentation thresholds (SUV 2.2, SUV 3, SUL 2.2, SUL 3, SUV_{BSA} 0.5 and SUV_{BSA} 0.7) for PFS at 6 months. TV using SUV 2.2, SUV 3, SUL 2.2 and SUL 3 all had an AUC of 0.867 (SE = 0.102, 95% CI = 0.666 – 1). Using SUV_{BSA} 0.5, AUC = 0.911 (SE = 0.080, 95% CI = 0.755 – 1) and for SUV_{BSA} 0.7, AUC = 0.889 (SE = 0.090, 95% CI = 0.712 – 1).

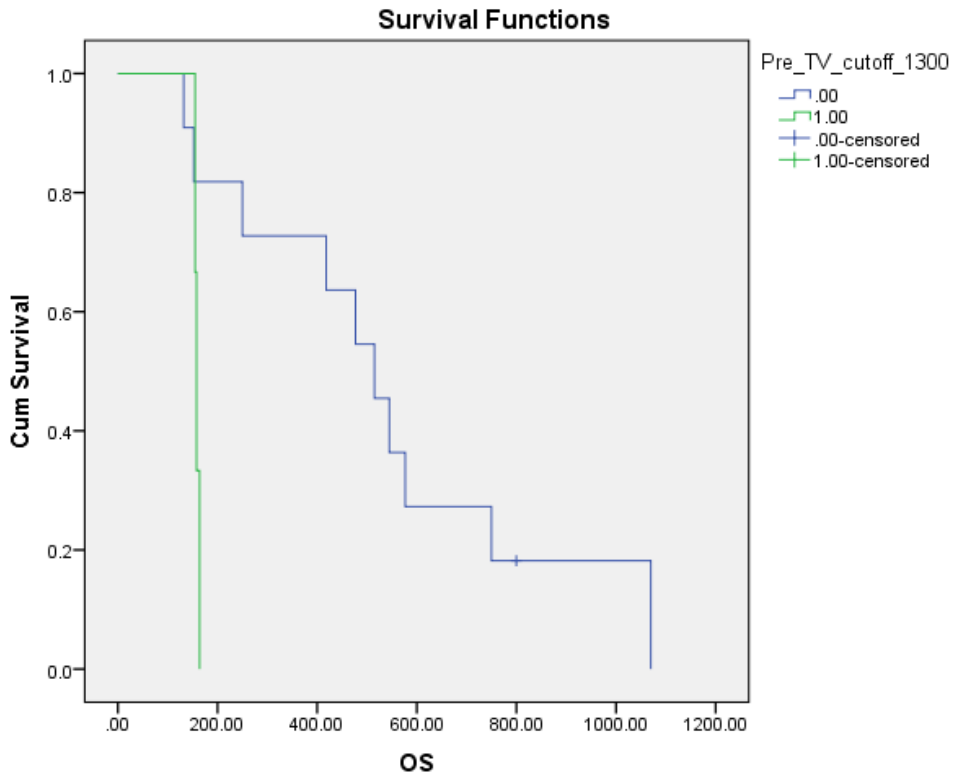


Figure 4.18: Kaplan-Meier Survival Curve for Pre-Therapy TV and OS

Survival curves are plotted for two groups of patients, those with pre- therapy TV > 1300ml (Mean PFS = 159, SE = 2.6, 95% CI = 154-164) and those < 1300ml TV (mean PFS = 541, SE = 97, 95% CI = 351-732) using a fixed 2.5 SUV threshold segmentation. Comparing the survival distributions using the log-rank method shows a significance of $p < 0.021$.

4.5.6) Predicting Survival in Patients with Mesothelioma using Tumour Volume and Total Lesion Glycolysis using Non-Fixed Threshold Segmentation

While fixed threshold methods had similar results in terms of predicting response, non-fixed threshold methods have shown less correlation between each other and are therefore more likely to show differences in predicting response. Correlations between the nine non-fixed threshold methods and OS and PFS for both pre- and post- therapy values and change and percentage change between them were calculated for both TV and TLG and are shown in Tables 4.24 and 4.25 respectively.

Segmentation Methods for TV	OS				PFS			
	Pre	Post	Change	% Chg	Pre	Post	Change	% Chg
Adapt SUV_{mean}	-0.599	-0.502	0.300	-0.017	-0.367	-0.276	0.208	-0.222
0.4 * Mean + 1	-0.624	-0.591	-0.167	0.100	-0.443	-0.443	-0.203	0.078
0.5 * Mean + 0.6	-0.368	-0.490	-0.450	0.183	-0.429	-0.359	0.012	0.525
0.5 * Mean + 1	-0.604	-0.641	-0.016	-0.084	-0.415	-0.448	-0.026	-0.164
PERCIST 2 S.D.	-0.649	-0.704	-0.416	0.189	-0.440	-0.489	-0.330	-0.160
PERCIST 3 S.D.	-0.630	-0.700	-0.272	0.270	-0.418	-0.479	-0.229	-0.235
Davis (2006)	-0.620	-0.686	-0.024	-0.073	-0.495	-0.510	0.035	-0.049
Nestle (2005)	-0.654	-0.687	-0.507	0.008	-0.452	-0.481	-0.369	-0.353
GRAB	-0.598	-0.699	0.015	0.424	-0.390	-0.463	-0.006	-0.063

Table 4.24: Correlation between TV with Non-Fixed Segmentation Thresholds and Survival

Pearson correlation coefficients (pccs) are shown for the correlation of TV, for each non-fixed threshold segmentation method, with OS and PFS for pre-therapy values, post-therapy values, and the change and percentage change (chg) from pre- to post- therapy values. For $p < 0.01$, $pcc > 0.662$ and $p < 0.05$, $pcc > 0.533$.

Segmentation Methods for TLG	OS				PFS			
	Pre	Post	Change	% Chg	Pre	Post	Change	% Chg
Adapt SUV_{mean}	-0.604	-0.494	0.306	0.040	-0.373	-0.267	0.218	-0.250
0.4 * Mean + 1	-0.629	-0.620	0.131	0.235	-0.420	-0.434	0.041	0.066
0.5 * Mean + 0.6	-0.562	-0.588	0.003	0.235	-0.427	-0.409	0.095	0.579
0.5 * Mean + 1	-0.616	-0.622	0.139	0.090	-0.406	-0.415	0.083	-0.146
PERCIST 2 S.D.	-0.643	-0.687	0.006	0.263	-0.424	-0.464	-0.023	-0.109
PERCIST 3 S.D.	-0.633	-0.683	0.055	0.341	-0.413	-0.455	0.015	-0.176
Davis (2006)	-0.619	-0.642	0.077	0.080	-0.454	-0.447	0.086	-0.076
Nestle (2005)	-0.644	-0.678	-0.051	0.221	-0.429	-0.462	-0.058	-0.246
GRAB	-0.612	-0.670	0.136	0.468	-0.396	-0.435	0.086	-0.042

Table 4.25: Correlation between TLG with Non-Fixed Segmentation Thresholds and Survival

Pearson correlation coefficients (pccs) are shown for the correlation of TLG, for each non-fixed threshold segmentation method, with OS and PFS for pre-therapy values, post-therapy values, and the change and percentage change (chg) from pre- to post- therapy values. For $p < 0.01$, $pcc > 0.662$ and $p < 0.05$, $pcc > 0.533$.

Different non-fixed threshold segmentation methods do not have an influence on TV and TLG in terms of predicting response. All nine methods generally have significance in predicting OS using

pre- and post- therapy values (for example, pcc ranging from -0.494 to -0.687 for pre-therapy TLG, p from ~0.07 to 0.007) but only show a weak correlation with PFS. There is no correlation with either OS or PFS using change or percentage change. These findings are supported by ROC analysis where good ROC curves are produced for six month OS, and to a lesser extent six month PFS, for all nine segmentation methods pre- and post- therapy TV and TLG (Table 4.26 and Table 4.27, respectively). Changes and percentage changes between pre- and post- therapy TV and TLG show poor accuracy in predicting six month OS and PFS. Some segmentation methods have slightly better ROC curves than others but the confidence intervals are large so it would be unfair to draw conclusions on which segmentation methods are more accurate. Results show that on this small cohort of mesothelioma patients the segmentation method makes no difference to the success or failure of TV and TLG in terms of predicting OS or PFS.

Segmentation Method for TV	**OS**				**PFS**			
	Pre	**Post**	**Change**	**% Chg**	**Pre**	**Post**	**Change**	**% Chg**
Adapt SUV_{mean}	0.889	0.711	0.244	0.378	0.875	0.708	0.313	0.438
0.4 * Mean + 1	0.889	0.867	0.600	0.467	0.813	0.792	0.521	0.396
0.5 * Mean + 0.6	0.844	0.844	0.578	0.422	0.750	0.729	0.542	0.479
0.5 * Mean + 1	0.867	0.911	0.689	0.644	0.875	0.854	0.583	0.583
PERCIST 2 S.D.	0.911	0.911	0.711	0.489	0.875	0.875	0.646	0.437
PERCIST 3 S.D.	0.889	0.911	0.667	0.511	0.875	0.875	0.604	0.458
Davis (2006)	0.933	0.867	0.511	0.489	0.938	0.833	0.396	0.417
Nestle (2005)	0.911	0.911	0.844	0.689	0.875	0.875	0.729	0.604
GRAB	0.889	0.911	0.578	0.422	0.875	0.875	0.458	0.354

Table 4.26: AUC for TV for Non-Fixed Segmentation Methods for Predicting Survival

Area under the curve (AUC) is shown for the ROC curves for TV using nine different non-fixed segmentation thresholds for predicting six month OS and PFS for pre- and post- therapy values and the change and percentage change between them.

Segmentation Method for TLG	OS				PFS			
	Pre	Post	Change	% Chg	Pre	Post	Change	% Chg
Adapt SUV_{mean}	0.889	0.711	0.289	0.311	0.875	0.729	0.292	0.313
0.4 * Mean + 1	0.911	0.889	0.467	0.444	0.875	0.812	0.396	0.375
0.5 * Mean + 0.6	0.889	0.844	0.511	0.333	0.833	0.771	0.479	0.354
0.5 * Mean + 1	0.867	0.867	0.489	0.511	0.875	0.833	0.396	0.458
PERCIST 2 S.D.	0.889	0.889	0.622	0.467	0.875	0.854	0.500	0.396
PERCIST 3 S.D.	0.889	0.889	0.467	0.467	0.875	0.854	0.375	0.396
Davis (2006)	0.889	0.867	0.400	0.444	0.875	0.833	0.313	0.375
Nestle (2005)	0.889	0.889	0.644	0.511	0.875	0.854	0.521	0.458
GRAB	0.889	0.867	0.467	0.356	0.875	0.833	0.375	0.292

Table 4.27: AUC for TLG for Non-Fixed Segmentation Methods for Predicting Survival

Area under the curve (AUC) is shown for the ROC curves for TLG using nine different non-fixed segmentation thresholds for predicting six month OS and PFS for pre- and post- therapy values and the change and percentage change between them.

It is worth noting that although segmentation methods produce similar ROC curves, the optimal thresholds are different for each method. This is not surprising given that different segmentation methods will produce smaller or larger TV based on the methodology used to segment images, however, it should be noted that the threshold for TV is dependent of the methods of segmentation. For example, when taking the optimal threshold for predicting six month OS using pre-therapy TV measures, a fixed 2.5 SUV method has an optimal threshold of ~1300ml compared to ~400ml for the adaptive SUV_{mean} method. Similar differences are seen between other segmentation methods too and this is an issue with using pre- and post- therapy values, not just for TV and TLG but also for SUV_{max} and any other parameter. In the case of TV and TLG, the optimal thresholds are likely to be dependent on the segmentation method, and similar to SUV_{max}, are most likely dependent on other factors such as the PET scanner used and imaging protocol. Therefore, even though the thresholds have been successful in identifying response in this cohort of patients, it is not viable to apply this to data from another institution. Although changes and percentage changes between pre- and post- therapy values have been unsuccessful at predicting

response, they can be more easily applied across institutions as they do not rely on definitive values.

4.5.7) Predicting Survival in Patients with Mesothelioma using Intensity Volume Histogram Parameters

All six IVH parameters implemented in PETTRA (discussed in 2.4.6, p121) were used to assess response in the cohort of mesothelioma patients using segmented TV from a fixed 2.5 SUV threshold. As these statistics are still related to intensity and TV it was thought that response would be similar to SUV_{max}, TV and TLG parameters. The parameter V_{10} was removed from analysis because it was found to be 100% in all but one image in which it was still over 99%. V_{10} is a representation of the volume having at least 10% of the SUV_{max} and in this cohort of patients it was clear that the almost all of the TV was over 10% of the SUV_{max} and therefore there would be no way to differentiate datasets into groups of reponders and non-responders. The other five IVH parameters were correlated with OS and PFS for pre- and post- therapy values and the change and percentage change between them (Table 4.28).

IVH Parameter	OS				PFS			
	Pre	Post	Change	% Chg	Pre	Post	Change	% Chg
V_{90}	0.539	0.528	0.214	-0.023	0.717	0.593	0.162	-0.126
V_{10-90}	-0.544	-0.528	-0.212	-0.212	-0.718	-0.593	-0.164	-0.164
I_{10}	-0.599	-0.385	0.315	0.360	-0.568	-0.516	0.080	-0.129
I_{90}	-0.559	-0.522	0.232	0.178	-0.487	-0.539	0.144	-0.005
I_{10-90}	-0.591	-0.378	0.316	0.366	-0.563	-0.514	0.074	-0.134

Table 4.28: Correlation between IVH Parameters and Survival

Pearson correlation coefficients (pccs) are shown for the correlation of five IVH parameters with OS and PFS for pre-therapy values, post-therapy values, and the change and percentage change (chg) from pre- to post- therapy values. For $p < 0.01$, $pcc > 0.662$ and $p < 0.05$, $pcc > 0.533$.

Correlations between IVH parameters and OS and PFS showed significant correlation with survival for most pre- and post- therapy values with no correlation between the change and percentage change between them. All parameters showed a negative correlation with OS and PFS except for V_{90} which has a positive correlation. This is not an expected correlation as V_{90} expresses the volume containing intensity values within 90% of the SUV_{max} so it would be expected a larger volume of intensity close to the SUV_{max} would result in a poor response. However, analysing the data it would appear that TVs with a high V_{90} have a lower SUV_{max} as there are similar lower SUVs in the TV. TVs with high SUV_{max} are unlikely to have other similarly high intensities in the TV.

Some of the IVH parameters are complicated to understand and relate to data than measures such as SUV_{max} and TV but they do show additional data which may be useful in response. With just 14 patients, it is hard to analyse the data to decipher the potential role they could play. In this case, SUV_{max}, SUV_{peak}, TV and TLG generally have better correlation with OS and PFS, making IVH parameters somewhat redundant, although that does not mean they should not be investigated in further studies with more patients where their role can be more clearly defined. Some IVH parameters produce very accurate ROC curves when using pre-therapy values (Table 4.29). While almost every pre-therapy TV has shown good correlation and ROC curves when compared with OS and PFS and the reasoning for this is clear, this is not the case with IVH with only some parameters correlating and not others, despite them being related to each other.

IVH Parameter	OS				PFS			
	Pre	Post	Change	% Chg	Pre	Post	Change	% Chg
V_{90}	0.022	0.100	0.556	0.444	0.000	0.115	0.604	0.542
V_{10-90}	1.000	0.900	0.444	0.444	0.990	0.885	0.396	0.396
I_{10}	0.756	0.533	0.289	0.311	0.875	0.646	0.208	0.250
I_{90}	0.811	0.667	0.278	0.333	0.885	0.688	0.188	0.250
I_{10-90}	0.756	0.533	0.289	0.311	0.875	0.646	0.229	0.250

Table 4.29: AUC for IVH Parameters for Predicting Survival

Area under the curve (AUC) is shown for the ROC curves for five IVH parameters for predicting six month OS and PFS for pre- and post- therapy values and the change and percentage change between them.

4.6) Conclusion of Response Analysis in Patients with Mesothelioma

In this study, 14 patients with mesothelioma underwent pre- and post- therapy PET/CT scans before and after therapy and several different methods of PET measurements were used to try and identify responders and non-responders. Clinical methods, such as visual analysis, and SUV_{max} and SUV_{peak} measures were used to try and identify response along with more novel measures such as TV and TLG, using a variety of different segmentation methods, and IVH parameters. With a small sample size (n=14), the study lacked statistical power, however, it would seem that pre- and post- therapy measures of SUV_{max}, SUV_{peak}, TV and TLG all provide good indication of survival in patients while visual analysis and the change and percentage change between SUV_{max}, SUV_{peak}, TV and TLG do not predict patient survival. IVH parameters had mixed results and while they do not show a good indication of response, they may be worthy of further investigation.

A comparison of different methodologies for obtaining SUV_{max} and SUV_{peak} parameters showed no real difference between them and their ability to predict OS or PFS. A similar comparison of segmentation methods also found no impact on TV or TLG values and their ability to predict OS

or PFS. This suggests that the segmentation method used for obtaining TV and TLG is unimportant, as long as one consistent method is used. However, further testing on a larger dataset would be needed before this can be concluded with confidence. Other novel methods of response including texture analysis and registered subtraction images are hoped to be used on this dataset and they may provide significant results. While many of the response methods tested in this analysis are related, these measures look at unrelated factors in PET images and may provide a different insight into the data.

It should be noted that this group of patients with mesothelioma had low survival times and a very poor prognosis, due to the aggressiveness and severity of the disease, meaning that there was no real response to treatment. This does not mean that the parameters could not potentially predict which patients were more likely to survive longer and therefore have some form of benefit from treatment, even if this was simply prolonging survival. However, it is likely that these factors mean that this data cannot be compared accurately to say a lymphoma dataset where a complete response is possible and often likely.

5) Response to Therapy in Patients with DLBCL

5.1) Introduction to Response to Therapy in Patients with DLBCL

5.1.1) Identifying Response to Therapy in DLBCL Patients

Having investigated response parameters in a group of patients with mesothelioma, similar methods were used to analyse response in a larger cohort of DLBCL patients, a disease with a higher survival rate due to more effective chemotherapy treatments. Once again, all patients in the study underwent pre- and post- therapy PET/CT scans and images were analysed using PETTRA software with the aid of an experienced clinician to exclude the inclusion of physiological areas of increased uptake in the segmentation of tumours volumes. Image analysis response parameters were compared with measures of survival to test their effectiveness.

5.1.2) Lymphoma

Lymphomas are a type of blood cancer which affects lymphocytes (white blood cells) causing them to behave abnormally. Lymphomas can be classified into two main groups depending on the presence or absence of the Reed-Sternberg cell, large cells which have a divided nucleus. Hodgkin's lymphoma (HL), named after its discoverer Thomas Hodgkin (Hodgkin, 1832), is characterised by the presence of the Reed-Sternberg cell, which is not found in other lymphomas known as non-Hodgkin's lymphomas (NHL). There are many different forms of NHL and since the 1980s they were classified differently using two classification systems: the Working Formulation adopted in the US, which classifies NHL based on its degree of aggressiveness, and the Kiel classification adopted in Europe and elsewhere, which classifies NHL based on cell morphology and B-cell and T-cell lineages (Lu, 2005). However, in 1994, the International Lymphoma Study Group developed a single classification system for international communication called the Revised European-American Classification of Lymphoid Neoplasms

(REAL) specifying clinically distinctive types of lymphoma with different prognostic groups (Harris *et al*., 1994). The WHO has since adopted this classification system with minimal modification and of the over 40 types of lymphoma there are six predominant subtypes which account for 80% of cases: DLBCL, follicular lymphoma, marginal zone B-cell lymphoma, small lymphocytic lymphoma, peripheral T-cell lymphoma and mantle cell lymphoma (Harris *et al*., 1997). The most common types of NHL are DLBCL and follicular lymphomas (Chan *et al*., 1997).

5.1.3) Diffuse Large B-Cell Lymphoma (DLBCL)

Diffuse large B-cell lymphoma (DLBCL) is an aggressive type of NHL which affects B-cells, a form of lymphocyte responsible for the production of antibodies (Dupas *et al*., 2013). In Europe and the US, the annual incidence of NHL is estimated to be 15-20 cases/100,000 and ~31% of them are DLBCL (Fisher and Fisher, 2004; Chan *et al*., 1997). Diagnosis of DLBCL is confirmed by examining tissue from a biopsy and the disease can be subdivided into four categories: (i) DLBCL – not otherwise specified, DLBCL with predominant extranodal location, (iii) large cell lymphomas of terminally differentiated B-cells and (iv) borderline cases (Martelli *et al*., 2013). The stage of the disease is distinguished using the Ann Arbor staging system which has four stages depending on the extent of spread, at stage I the disease is located in a single region whereas in stage IV the disease has spread to the extra-lymphatic organs such as the liver, spleen and bone barrow (Carbone *et al*., 1971). The cause of DLBCL is not known but usually arises from normal B-cells mutating although it can also be the result of a transformation from other types of less aggressive lymphomas or leukaemia (Martelli *et al*., 2013).

5.1.4) Treatment of DLBCL

DLBCL has become a curable lymphoma since the 1970s following the introduction of the chemotherapy regime cyclophosphamide, doxorubicin, vincristine, and prednisolone (CHOP)

(McKelvey *et al*., 1976). This chemotherapy regime has been improved further since, with the addition of monoclonal anti-CD20 antibody rituximab (R-CHOP), which has increased the event-free survival (EFS) and OS compared to the CHOP regime (Coiffier *et al*., 2002; Feugier *et al*., 2005; Coiffier *et al*., 2010; Pfreundschuh *et al*., 2011). R-CHOP chemotherapy now means that patients with DLBCL can now expect 5-year PFS rates of 55% for patients >60 and 75% for patients <60 (Friedberg *et al*., 2008). Despite improvements in chemotherapy with R-CHOP, 30-40% of patients are still not cured and salvage treatment appears inadequate in the majority of these patients so a strategy to improve patient outcome is to intensify first-line treatment using high dose therapy and/or autologous stem cell transplantation (ASCT) which has been efficient in curing nearly half of all patients who have resistant or relapsed NHL (Zinzani *et al*., 2011). High dose chemotherapy and ASCT are less effective after R-CHOP so a 'response-adapted therapy' strategy where non-responding patients change therapy regimes mid-treatment is a potential solution (Mikhaeel, 2006; Kasamon *et al*., 2009; Mikhaeel, 2009). In order to do this a reliable method of identifying responding and non-responding patients is needed early in the course of treatment.

One method of trying to separate responding and non-responding patients is to risk stratify patients before treatment with the International Prognostic Index (IPI). The IPI is a model derived from tests on over 4000 patients to determine factors that are associated with prognosis. The IPI uses clinical parameters such as disease stage, lactate dehydrogenase level, ECOG performance score and extra-nodal disease (International NHL Prognostic Factors Project, 1993). There is also a revised version adjusted for age, called the age adjusted IPI (aaIPI), and a version specifically for DLBCL patients treated using R-CHOP (R-IPI) (Sehn *et al*., 2007). Patients divided into risk categories according to their IPI have been shown to have significantly different survival times to suggesting that it is a useful tool to predict prognosis (Ziepert *et al*., 2010; Martelli *et al*., 2013). However, although the IPI is valuable for stratification groups of patients in clinical trials, the

prediction of outcome is more variable for an individual patient so using the index to tailor treatment regimes for the individual patient may be flawed (Sweetenham, 2005).

5.1.5) Role of ^{18}F-FDG PET/CT in Management of DLBCL

^{18}FDG-PET has been shown to accurately detect disease in DLBCL and other types of NHL (Elstrom *et al*., 2003), and, along with R-IPI, pre-therapy SUV_{max} has been shown to predict patients with longer PFS (Chihara *et al*., 2011; Oh *et al*., 2012; Miyazaki *et al*., 2013). To assess response to therapy with a view to changing treatment, an interim-PET scan during treatment is considered to be a valuable tool (Mikhaeel *et al*., 2005). Interim-PET scans have been shown to predict EFS, PFS and OS strongly in a number of studies, often much better than the IPI (Jerusalem *et al*., 1999; Spaepen *et al*., 2002; Haioun *et al*., 2005; Mikhaeel *et al*., 2005; Dupuis *et al*., 2009; Safar *et al*., 2012). However, this is not always the case and some studies have recently reported that patients with DLBCL have similar PFS for patients with PET-positive scans (suggesting active disease) and PET-negative scans (suggesting no disease) scans after two to four cycles of chemotherapy with poor positive predictive values of outcome (Moskowitz *et al*., 2010; Cashen *et al*., 2011; Pregno *et al*., 2012). These conflicting results may be down to different visual criteria being used (Terasawa *et al*., 2009), a lack of inter-observer reproducibility in interpreting PET scans (Horning *et al*., 2010) or a combination of both.

The visual interpretation of interim/mid-treatment PET scans for identifying response in DLBCL patients is a controversial topic and international workshops have constantly revisited and revised visual response criteria (Meignan *et al*., 2009; Meignan *et al*., 2010; Meignan *et al*., 2012). When using visual analysis, categorising patients into two groups of PET negative patients/responders and PET positive patients/non-responders is often done to try and predict outcome (Spaepen *et al*., 2002; Haioun *et al*., 2005). However, a study at Guy's and St Thomas's Hospitals, showed that dividing patients into three categories of PET negative, PET positive and minimal residual

uptake (MRU) could be more beneficial with five-year PFS's of 88.8% for PET negative patients, 59.3% for those with MRU and 16.2% for PET positive patients (Mikhaeel *et al*., 2005). In an attempt to improve on the three category system, a 5-point scoring system has been proposed which scores residual uptake compared to the uptake in normal reference regions in the body e.g. mediastinal blood pool and liver (Mikhaeel, 2009; Barrington *et al*., 2010). The 5-point scoring system, shown in Table 5.1, has been recommended as the most suitable criteria for assessing response in lymphoma patients during the latest International Workshop criteria (Meignan *et al*., 2010; Meignan *et al*., 2012).

Score	Category	Description
0	Complete Metabolic Response (CMR)	No Uptake
1	Minimal Residual Uptake 1	Uptake <= Mediastinum
2	Minimal Residual Uptake 2	Uptake > Mediastinum but <= Liver
3	Residual Lymphoma	Uptake > Liver
4	Progressive Disease	New lesion(s) likely to be lymphoma

Table 5.1: Five-Point Scoring System for Patients with DLBCL

Five point scoring system for visual analysis of mid-treatment PET scans for patients with DLBCL (Adapted from Mikhaeel, 2009).

Despite these steps forward in terms of defining visual criteria, inter-observer reproducibility is still questionable. While the criteria proved to have good inter-observer reproducibility at four European centres reporting scans in patients with HL, with Kappa agreements between pairs of experts of 0.85 and 0.79 when using the liver and mediastinum as reference backgrounds respectively (Barrington *et al*., 2010), reproducibility between experts is not as strong for DLBCL patients (Horning *et al*., 2010). This has led to researchers investigating more 'quantitative' approaches of assessing response such as the change in SUV_{max} between pre- and post- therapy scans (Δ SUV_{max}) (Lin *et al*., 2007; Meignan *et al*., 2010, Meignan *et al*., 2012). Despite the issues with visual criteria, the majority of studies have still shown that interim-PET using visual

analysis can predict OS or PFS (Jerusalem *et al*., 1999; Spaepen *et al*., 2002; Haioun *et al*., 2005; Mikhaeel *et al*., 2005; Dupuis *et al*., 2009; Safar *et al*., 2012), and one study which altered treatment based on visual assessment of the interim PET scan and showed favourable results for modifying treatment (Kasamon *et al*., 2009).

5.1.6) Use of Semi-Quantitative Methods for the Management of DLBCL

A more quantitative approach to response assessment reduces the subjectivity and observer variability compared to visual analysis. A quantitative approach after two cycles of chemotherapy has been reported to be better than visual assessment in terms of predicting treatment outcome, using a change in SUV_{max} (Δ SUV_{max}) of ~66% between pre- and post- therapy values (Torizuka *et al*., 2003; Lin *et al*., 2007; Meignan *et al*., 2009; Casanovas *et al*., 2011; Safar *et al*., 2012). After four cycles of chemotherapy a slightly higher cut-off for the Δ SUV_{max} of 70-73% has also shown to successfully predict outcome (Itti *et al*., 2009; Casanovas *et al*., 2011). Using Δ SUV_{max} appears to minimise the risk of false positive results obtained using visual analysis with a higher positive predictive value of over 50% (Itti *et al*., 2009; Casanovas, 2012). A caveat to wider application of these methods is that careful standardisation of PET methods is required for accurate and reproducible measurement of quantitative parameters (Boellaard *et al*., 2009).

As well as SUV_{max} measurements, studies have begun to investigate volumetric measures such as metabolic tumour volume (MTV) and total lesion glycolysis (TLG) for predicting response in DLBCL patients with varying results (Kim *et al*., 2012; Song *et al*., 2012). Pre-therapy TLG has been shown to predict 2-year PFS and OS better than pre-therapy SUV_{max} (Kim *et al*., 2012), while patients with high pre-therapy MTV have been shown to have lower PFS and OS (Song *et al*., 2011; Song *et al*., 2012). Different segmentation methods were used to obtain TV and TLG in these studies with thresholds based on a percentage of SUV_{max} (Kim *et al*., 2012), and contouring of uptake >2.5 SUV (Song *et al*., 2011; Song *et al*., 2012), both being implemented. These results

show that along with Δ SUV_{max}, Δ TV and Δ TLG may prove to be useful measures of predicting response in patients with DLBCL.

5.1.7) Aim of Analysis

The main aim of this analysis was to determine if a quantitative method would better discriminate patients into groups of good responders and poor responders than existing methods like IPI and visual assessment. The parameters investigated as prognostic factors were the IPI score of the patient, visual analysis using the five-point scoring system, SUV_{max}, TV and TLG. Pre-therapy, post-therapy and the change during treatment of SUV_{max}, TV and TLG were assessed. SUV_{peak} was ommited as the difference between SUV_{peak} and SUV_{max} was found to be negligible in terms of assessing response on patients with mesothelioma (see 4.5.4).

5.2) Patients and Scanning

5.2.1) Patient Eligibility and Treatment

85 patients with a *de novo* diagnosis of DLBCL and no concurrent low grade lymphoma underwent both pre- and post- therapy PET/CT scans before and after two cycles of R-CHOP or Rituximab with Cyclophosphamide, Etoposide, Prednisolone, and Vincristine (R-CEOP) chemotherapy, an altered regime, similar to R-CHOP, showing excellent outcome in patients with a contraindication to anthracyclines (Moccia *et al*., 2009). All patients had assessable disease on the pre-therapy PET scan and had a minimum follow-up of at least 12 months.

5.2.2) PET/CT Scanning

Pre- and post- therapy PET/CT scans were acquired at baseline and after two cycles of chemotherapy for 85 patients between April 2005 and February 2011. All scans were acquired at

the PET Imaging Centre in St Thomas' Hospital on either a GE Discovery ST or GE Discovery VCT PET scanner (Waukesha, WI). For the PET and CT scan components of the PET/CT study, patients were scanned during free breathing and CT scans were low dose non-contrast enhanced scans. For the 170 whole body PET scans over the 85 datasets, image dimensions were 128 x 128 x 179, 223, 263, 267 or 311 (with voxel sizes of either 5.47mm x 5.47mm x 3.27mm – on 39 images – or 4.69mm x 4.69mm x 3.27mm – on 131 images). All PET images analysed were attenuation corrected using a smoothed CT dataset. Administered FDG activity ranged from 250 to 399MBq (mean: 346MBq; S.D. 24.76). The difference in administered activities between pre- and post- therapy scans ranged from 0-96MBq (mean: 27.6MBq; S.D. 21.66). The time between administration of FDG and the start of the scan ranged from 70-148min (mean: 97min; S.D. 13.96), with 45/85 post-therapy scans within 10min of the pre-therapy scans and 68/85 post-therapy scans within 20min of the pre-therapy scans.

5.2.3) Patient Characteristics

Patients had a mean age of 57 (range 25-86, S.D. 14.59) and 47% were male (40 male, 45 female). Patients weighed between 37-121kg (mean: 71kg; S.D. 17.06) and there was a mean decrease in weight of 3.6kg between pre- and post- therapy scans (range: 0-14kg), most likely due to the effects of the disease and the chemotherapy. Patient height ranged from 149-192cm (mean: 169cm; S.D. 10.27) with a mean change in height between scans of 1.1cm (range: 0-10cm).

5.2.4) Study End Point

Overall survival (OS) and progression free survival (PFS) were the outcome measures used to assess response, as for the mesothelioma dataset and other studies investigating patients with DLBCL (Jerusalem *et al*., 1999; Mikhaeel *et al*., 2005; Safar *et al*., 2012; Kim *et al*., 2012).

5.3) Data Analysis – Response Measures

5.3.1) Visual Analysis

Visual analysis of pre- and post- therapy PET/CT studies was performed by a consultant physician with 20 years of experience of PET. Response was assessed using both PET and CT, viewed on HERMES Hybrid Viewer workstations (Nuclear Diagnostics AB, Stockholm, Sweden). Visual assessments of response were scored based on the Deauville criteria (Table 5.1). SUV_{max} values were also measured by the consultant to confirm the visual assessment of uptake as higher than the mediastinum and/or liver when uptake was present in lesions seen on the pre-therapy scan. Scans were viewed scaled to a SUV_{max} of 10 normalised for injected activity and body weight.

5.3.2) SUV_{max}

SUV_{max} was taken as the maximum SUV within the entire TV on the HERMES Hybrid Viewer software by the consultant physician and on PETTRA software within the segmented TV, the two being identical for all values.

5.3.3) Tumour Volume Segmentation

Due to its simplicity and successful application on mesothelioma data and its previous use in patients with DLBCL (Song *et al*., 2011; Song *et al*., 2012), a fixed 2.5 SUV threshold segmentation method was chosen to produce TV on all the images. The consultant physician for reviewed the TV to ensure that areas with increased physiological uptake were not included. While it would have been interesting and beneficial to test different segmentation methods on a larger dataset, unfortunately, this was considered too time consuming to be consider in this study.

5.3.4) Tumour Volume and Total Lesion Glycolysis

TV in a PET scan was taken as the total of all segmented areas of disease. TV was calculated as the volume of each voxel multiplied by the number of segmentation voxels, given in ml. TLG was calculated as the product of TV and SUV, given in ml*SUV.

5.4) Segmentation of Patients with DLBCL

A fixed 2.5 SUV threshold region growing segmentation method was used to delineate disease in all the images of patients with DLBCL. A consultant physician confirmed that all areas contained within the TV were due to lymphomatous involvement in her opinion rather than physiological uptake or another process that can cause increased FDG uptake such as infection or inflammation. Head and neck and brain images were also used as part of the disease segmentation if extra views were obtained in patients in addition to a half-body scan. Disease was segmented on 122 whole body images, 20 head and neck images and 2 brain images on pre- and post- therapy scans over the 85 patient dataset. Disease was always present on one of the pre-therapy images (whether whole body, head and neck or brain), however, 32 patients had no disease with uptake >2.5 SUV on post-therapy images and therefore no TV (and therefore TLG) was computed for these patients with a CMR. All the post-therapy images with no TV or TLG were also considered to have no disease of any form by the consultant physician when performing visual analysis.

Over the 85 patient dataset, 115 images needed no restrictions to stop the fixed 2.5 SUV threshold segmentation ‘spilling’ over into physiological areas that did not represent lymphomatous disease. In the rest of the dataset a total of 95 VOIs were used to restrict the segemented MTV from extending into areas where there was no disease or to remove physiological uptake from being incorrectly included in the MTV such as urinary uptake or on one occasion to avoid ‘double-counting’ an area of tumour which was present on both the body and head and neck images. The local views usually contain areas that overlap with the body images to avoid missing

any areas of disease when staging and assessing response. The areas of physiological uptake referred to included the brain, tonsils, pharynx, tongue, lung, heart, spleen, liver, kidney, bowel, bladder and bone marrow (Table 5.2). Over the 85 patients, there were a total of 939 segmented areas of disease, 840 on pre-therapy scans and 99 on post-therapy scans. Pre-therapy images had a mean of 9.77 areas of disease (range: 1-59, S.D.: 11.03) while post-therapy images had a mean of just 1.15 areas (range: 0-9, S.D.: 11.03). The mean TV was 887.54ml (median: 391.38ml, range: 1.50-4616ml, S.D.: 1073) in pre-therapy images and 34.88ml (median: 0.65ml, range: 0-1609ml, S.D.: 188) in post-therapy images.

Physiological Uptake Area	Restricted	Removed	Total	VOIs Needed
Brain	8	1	9	12
Tonsils	3	0	3	3
Pharynx	5	0	5	5
Tongue	4	0	4	4
Lung	1	0	1	4
Heart	2	2	4	4
Spleen	0	1	1	1
Liver	4	1	5	5
Kidney	8	11	19	23
Bowel	5	1	6	8
Bladder	14	5	19	19
Bone Marrow	5	0	5	5
ALL	59	22	81	93

Table 5.2: Physiological Uptake Areas Segmented using a Fixed 2.5 SUV Threshold

Over the dataset, 81 areas of physiological uptake were segmented with disease using a fixed 2.5 SUV region growing segmentation algorithm. Physiological uptake was either removed from the segmented area of disease or restricted to a particular volume so it did not 'spill' into areas of physiological uptake. VOIs Needed refers to the number of cubic VOIs that were needed to remove the physiological uptake.

The segmentations in the DLBCL dataset were more complex than in the previous mesothelioma dataset, partially due to the substantial increase in patients, but also because DLBCL is a disease which can affect many different areas of anatomy with a widespread distribution often throughout the body in comparison to mesothelioma, which is more often restricted to the thorax and sometimes a hemi-thorax. As Table 5.2 shows, areas of physiological uptake often incorrectly segmented as lymphoma are situated in widely separated parts of the body, with the most common being the brain, kidneys and bladder. This is to be expected as they are often very intense areas of physiological uptake in areas where lymphatic tumours are likely to present. An example of the segmented areas of disease in DLBCL patients is shown in Figure 5.1 where six pre-therapy scans are shown with segmentations.

5.5) Predicting Survival in Patients with DLBCL

5.5.1) Introduction to Predicting Survival in Patients with DLBCL

Response measures for visual analysis, SUV_{max}, TV and TLG were used to predict PFS and OS over the 85 patients. Ann Arbor staging and the IPI were also included in the as comparators. Pearson correlation coefficients (pcc), ROC curves and Kaplan-Meier survival curves were used for the statistical analysis to determine if there was an association between response measures and PFS/OS (described in more detail in 4.5.2). Statistical analysis was done using statistical software package SPSS Statistics Version 20 (IBM, Armonk, NY, USA).

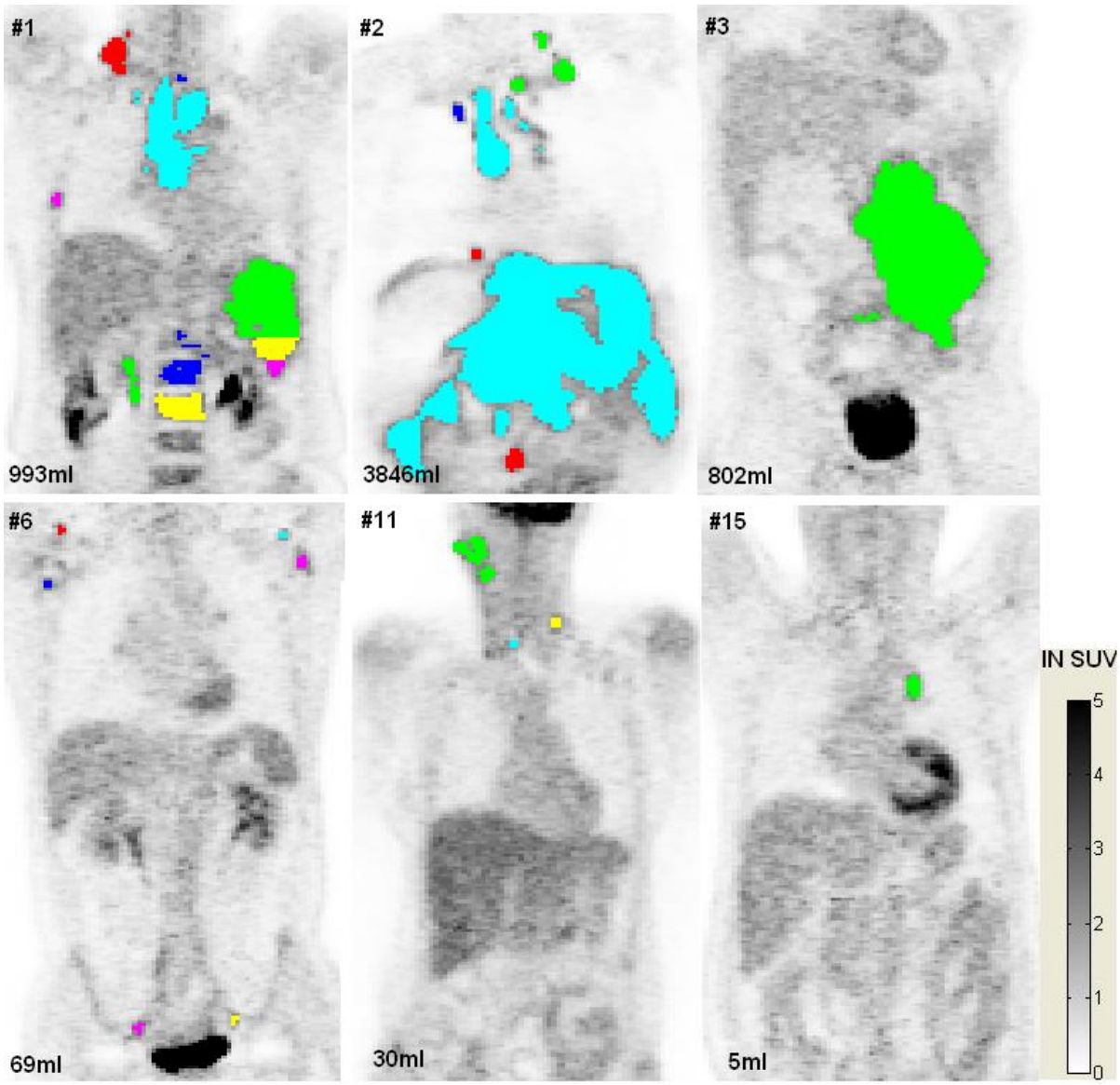


Figure 5.1: Segmentations of 6 Pre-Therapy Images of Patients with DLBCL

Six pre-therapy images of patients with DLBCL, all of which were within the first 20 patients in the dataset, show the difference in TV and regions of disease in different patients. While the first three scans have relatively large TVs with disease spread throughout the body, datasets 6, 11 and 15 have smaller TVs with disease in more specific regions. Disease is focused in the infraclavicular, axillary and pectoral lymph nodes and iliac, inguinal and femoral lymph nodes in dataset 6 (inguinal and femoral nodes not shown on this slice), the cervical, supraclavicular occipital and pre-auricular lymph nodes in dataset 11 (occipital and pre-auricular lymph nodes not shown on this slice) and localised to the mediastinum in dataset 15. Different colours indicate different VOIs of the segmentation of disease.

5.5.2) PFS and OS in Patients with DLBCL

Patients with DLBCL had much longer survival times in comparison to the mesothelioma patients with a mean PFS of 1126 days and a mean OS of 1272 days (Table 5.3). There are longer survival times for many reasons, mainly because mesothelioma is a particularly aggressive disease in comparison to DLBCL, presents later and R-CHOP and R-CEOP often cure DLBCL patients resulting in long survival times. This means a larger study population is required to test associations with survival in DLBCL than mesothelioma because there will be fewer events (disease progression, relapse or death). In the mesothelioma dataset just one patient out of fourteen was still alive, in the patients in this dataset just 23/85 patients had died. However, there was a good duration of follow up for the majority of patients in the study so for most patients, the fact they were still alive would indicate a good response to treatment. Figure 5.2 shows the survival curves for both PFS and OS, with the longer survival times compared to mesothelioma curves clear to see (Figure 4.15).

PFS	
Mean	1126
Median	1176
S.D.	697
Min	63
Max	2485

OS	
Mean	1272
Median	1273
S.D.	610
Min	106
Max	2485

Table 5.3: PFS and OS for DLBCL Patients

Of the 85 patients, 23 were deceased, 8 were alive but had confirmed disease progression and 54 were still alive and had no confirmed disease progression.

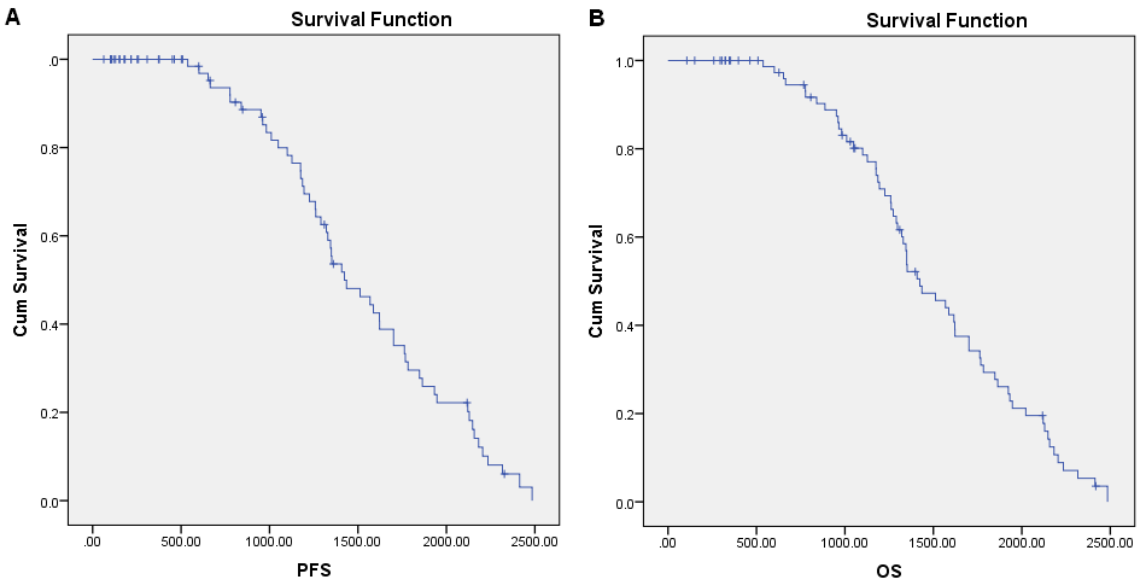


Figure 5.2: Kaplan-Meier Survival Curves for all DLBCL Patients

Survival curves for (A) PFS and (B) OS for all 85 DLBCL patients. OS and PFS are in days.

5.5.3) Predicting Survival in DLBCL Patients using Staging, IPI and Visual Analysis

Visual analysis was done using the Deauville criteria five point scoring system (Table 5.1) and the breakdown of patients for each score is shown in Table 5.4 along with the Ann Arbor staging and IPI which can both be used as prognosis factors for predicting response. Most patients were diagnosed with advanced stage DLBCL with an Ann Arbor staging of IV, while the IPI was distributed more evenly.

The correlation between Ann Arbor stage, IPI and Deauville criteria with OS and PFS shows good correlation for all three methods with higher correlations for Ann Arbor staging and Deauville criteria with correlations of -0.467 and -0.300 for PFS ($p < 0.006$), and -0.289 for IPI ($p < 0.008$). Negative correlations mean that the higher the Ann Arbor, IPI and Deauville score, the

lower the PFS and OS which is to be expected as the higher the score for all three criteria the worse the prognosis. .

Ann Arbor Stage	
Stage	# Patients
I	9
II	16
III	10
IV	50
Total	85

IPI	
Score	# Patients
0	8
1	19
2	17
3	17
4	17
5	7
Total	85

Deauville Criteria	
Score	# Patients
1	25
2	8
3	18
4	24
5	10
Total	85

Table 5.4: Ann Arbor Staging, IPI and Visual Deauville Criteria for Patients with DLBCL

The number of patients (# patients) in each category according to Ann Arbor staging, IPI and Deauville scoring systems is displayed.

Pearson Correlation Coefficients			
	Ann Arbor	IPI	Deauville
PFS	-0.467	-0.289	-0.300
OS	-0.433	-0.271	-0.297

Table 5.5: Correlation between Ann Arbor Staging, IPI and Deauville Score and Survival

Pearson correlation coefficients (pccs) are shown for the correlation of Ann Arbor staging, IPI and Deauville score with OS and PFS. p between 0.012 and 0.0001 for all correlations.

Mean PFS for each Ann Arbor stage, IPI score and Deauville score shows that all the methods struggle to separate patients into groups with significantly different PFS and therefore struggle to predict patient survival. Ann Arbor staging produces clear differences in survival between patients in stages I/II (mean PFS of 1747 for stage I and 1769 for stage II) compared to those in stages III/IV (mean PFS of 1473 for stage III and 1329 for stage IV). The IPI shows a general trend of higher scores correlating with shorter PFS. However, although a score of 0 results in a

much longer average PFS (mean = 1854), the mean PFS is similar for IPI scores 1-4 and patients with a score of 3 actually have a lower mean PFS estimate than those with a score of 4 or 5. Similar PFS estimates are seen with Deauville scores where patients with a score of 3 have longer mean PFS estimates than patients with a score of 1 or 2.

Mean PFS Times for Ann Arbor Staging, IPI and Deauville Scores									
Score	**Ann Arbor Stage**			**IPI**			**Deauville Score**		
	Estimate	**SE**	**95% CI**	**Estimate**	**SE**	**95% CI**	**Estimate**	**SE**	**95% CI**
0	N/A			1854	102	1654-2053	N/A		
1	1747	206	1343-2152	1598	151	1302-1894	1608	116	1382-1835
2	1769	120	1534-2003	1475	179	1124-1826	1462	167	1135-1790
3	1473	211	1059-1887	1275	174	934-1617	1643	148	1354-1933
4	1329	82	1169-1489	1455	115	1230-1681	1346	134	1083-1609
5	N/A			1400	77	1249-1552	1379	296	798-1960

Table 5.6: Mean PFS for Ann Arbor Staging, IPI and Deauville Score

Mean PFS estimates with standard error (SE) and 95% confidence intervals (95% CI), using Kaplan-Meier analysis, are shown according to Ann Arbor stage, IPI score and Deauville score for all 85 DLBCL patients.

ROC analysis further confirms these findings with curves for all three scoring systems running close to the reference line of taking random guesses to be considered useful at predicting 5-year PFS. Similar curves were observed for predicting 5-year OS too. 5-year PFS and OS was chosen as a binary threshold for ROC analysis because a similar time frame has been used in many other studies investigating survival in DLBCL patients (Mikhaeel *et al*., 2005; Dupuis *et al*., 2009). All the statistical results suggest that Ann Arbor stage, IPI score and Deauville score are not suitable methods of predicting survival in DLBCL patients.

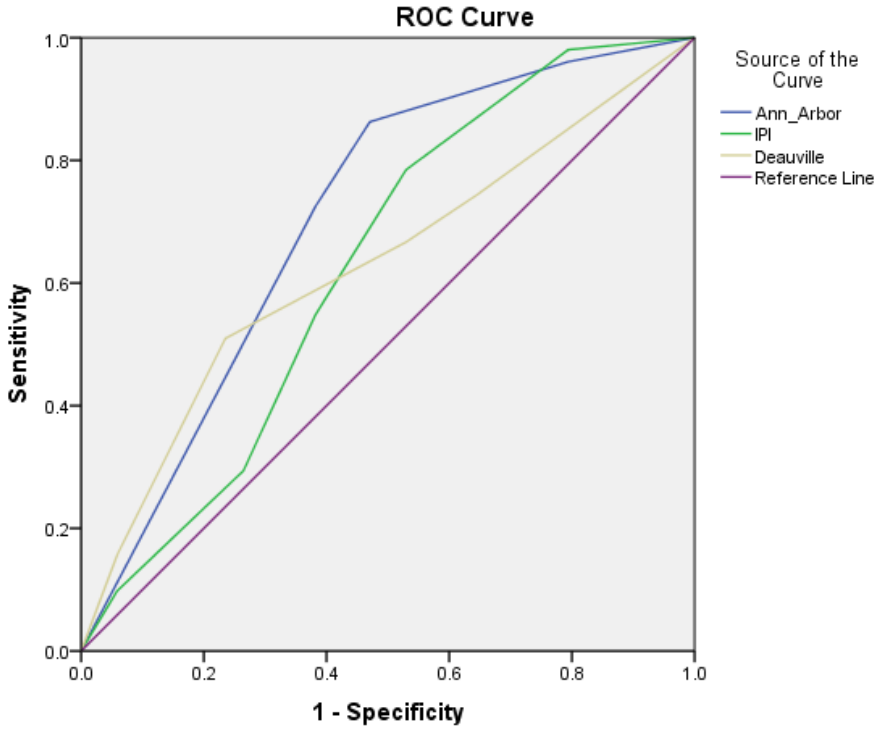


Figure 5.3: ROC Curve for Ann Arbor Staging, IPI and Deauville Score for 5-year PFS

ROC curves are plotted for Ann Arbor staging, IPI and Deauville scores for 5-year PFS. All curves show limited ability to predict PFS. Ann Arbor staging has an area under the curve (AUC) of 0.706 (Standard Error (SE) = 0.60, 95% Confidence Intervals (CI) of 0.587 – 0.824), IPI an AUC of 0.628 (SE = 0.65, 95% CI = 0.501 – 0.755) and Deauville score an AUC of 0.627 (SE = 0.61, 95% CI = 0.508 – 0.747).

5.5.4) Predicting Survival in DLBCL Patients using SUV_{max}, TV and TLG

Semi-quantitative response measures were taken pre- and post- therapy so can be used to predict response using either of these values or using the change or % change between them. TV and TLG both were calculated using the fixed 2.5 SUV segmentation method discussed in 5.4. Correlations between each response measure and survival are shown in Table 5.7. The best correlation with response was for the post-therapy SUV_{max} (pcc = -0.407, $\rho < 0.001$ for PFS) and the % change in SUV_{max} (pcc = -0.389, $\rho < 0.001$ for PFS). Apart from pre-therapy SUV_{max}, all

measures showed some degree of small or moderate correlation with both PFS and OS although no correlations could be described as strong. Interestingly, the change in TV and TLG were correlated positively rather than negatively with PFS and OS as would have been expected, as were the mesothelioma results, perhaps for similar reasons (more likelihood of greater change in patients with initially high TV and TLG who are also less likely to respond).

Response Parameter	PFS				OS			
	Pre	Post	Change	% Chg	Pre	Post	Change	% Chg
SUV_{max}	0.016	-0.407	-0.149	-0.389	-0.093	-0.405	-0.046	-0.357
TV	-0.228	-0.231	0.199	-0.216	-0.201	-0.280	0.161	-0.246
TLG	-0.276	-0.228	0.244	-0.244	-0.267	-0.278	0.224	-0.280

Table 5.7: Correlation between SUV_{max}, TV and TLG with Survival

Pearson correlation coefficients (pccs) are shown for the correlation of SUV_{max}, TV and TLG with OS and PFS for pre-therapy values, post-therapy values, and the change and percentage change (chg) from pre- to post- therapy values. For $p < 0.05$, pcc > 0.217 and $p < 0.01$, pcc > 0.284.

ROC analysis of all the parameters showed little success at predicting 5-year PFS and OS with AUC between 0.408 and 0.646 (Table 5.8). The ROC curves for pre-therapy SUV_{max}, TV and TLG and the % change between the pre- and post- therapt parameters for 5-year PFS are shown to illustrate this (Figure 5.4). However, ROC analysis is designed to divide patients into two groups of responders and non-responders using just one threshold. This may not always be the best way of dividing patients or separating patients who respond from those who don't. Many studies have however reported that a change in SUV_{max} of -66% has been a good threshold of separating responders and non-responders after two cycles of chemotherapy (Torizuka *et al*., 2003; Lin *et al*., 2007; Meignan *et al*., 2009; Casanovas *et al*., 2011; Safar *et al*., 2012). If applied to this dataset, this does produce survival curves with patients separated with different estimates of survival of >800 days but only 11 out of the 85 patients have a change of SUV_{max} less than -66% (Figure 5.5).

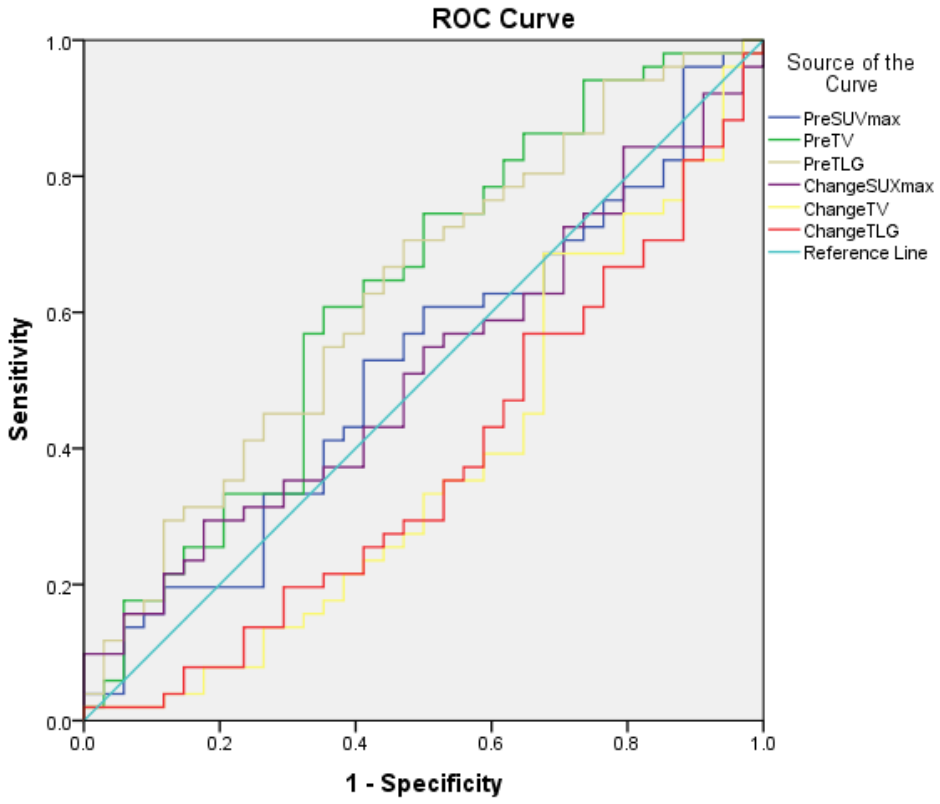


Figure 5.4: ROC Curve for SUV_{max}, TV and TLG for Predicting 5-year PFS

ROC curves are plotted for pre-therapy SUV_{max}, TV and TLG and the percentage change between them for predicting 5-years PFS. All curves show limited ability to predict PFS. Pre-therapy SUV_{max} has an area under the curve (AUC) of 0.521 (Standard Error (SE) = 0.64, 95% Confidence Intervals (CI) of 0.395 – 0.646), pre-therapy TV an AUC of 0.628 (SE = 0.63, 95% CI = 0.504 – 0.752), pre-therapy TLG an AUC of 0.630 (SE = 0.62, 95% CI = 0.508 – 0.751). Percentage change between pre-therapy and post-therapy SUV_{max} has an AUC of 0.515 (SE = 0.63, 95% CI = 0.391 – 0.639), change for TV an AUC of 0.378 (SE = 0.64, 95% CI = 0.253 – 0.503) and change for TLG an AUC of 0.376 (SE = 0.62, 95% CI = 0.254 – 0.498).

Response Parameter	5 Year PFS				5 Year OS			
	Pre	Post	Change	% Chg	Pre	Post	Change	% Chg
SUV_{max}	0.521	0.628	0.515	0.641	0.552	0.600	0.475	0.598
TV	0.628	0.646	0.378	0.628	0.596	0.612	0.410	0.599
TLG	0.630	0.645	0.376	0.627	0.599	0.611	0.408	0.594

Table 5.8: AUC for SUV_{max}, TV and TLG for Predicting 5 Year Survival

Area under the curve (AUC) is shown for the ROC curves for SUV_{max}, TV and TLG for predicting five year PFS and OS for pre- and post- therapy values and the change and percentage change between them. Standard Error is between 0.60 and 0.65 for all values.

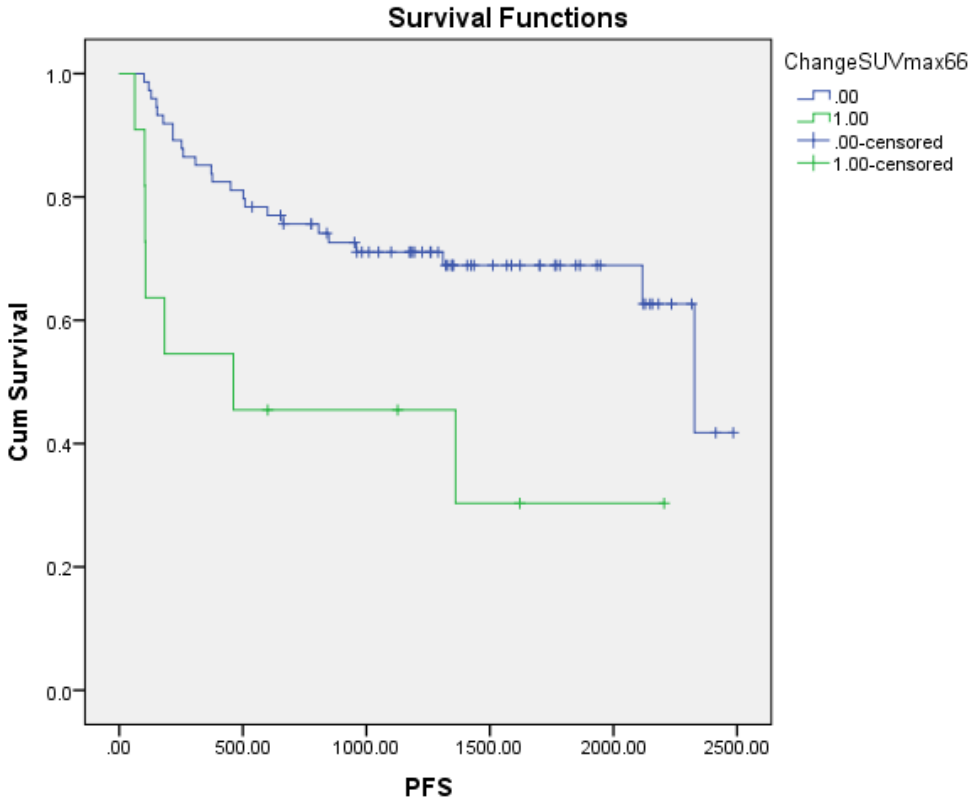


Figure 5.5: Kaplan-Meier Survival Curve for 66% Change in SUV_{max} and PFS

Survival curves are plotted for two groups of patients, those with change in SUV_{max} from -100% to -66% (mean PFS = 967, SE = 285, 95% CI = 409-1525 and median PFS = 462, SE = 565, 95% CI = 0-1570) and those with change in SUV_{max} from 100% to -65.9% (mean PFS = 1800, SE = 113, 95% CI = 1578-2022 and median PFS = 2328, SE = 180, 95% CI = 1975-2681). Comparing the survival distributions using the log-rank method shows a significance of $p < 0.006$.

The results suggest that there is some correlation between SUV_{max}, TV and TLG with PFS, suggesting they may have some use in predicting outcome but using ROC analysis and Kaplan-Meier curves to examine associations between these parameters and response showed no real benefit in using these methods to predict response. Different approaches to separating patients into prognostic groups may result in more promising results which are more suitable for use in patients with DLBCL. Dividing patients into three groups (tertiles) rather than two appears like it may produce statistically significant results. Statistical analysis on this dataset is still ongoing, as is the investigation of other exploratory measures for response including texture analysis and subtraction images from registered pre- and post- therapy scans.

5.6) Conclusion to Response to Therapy in Patients with DLBCL

In this chapter, the ability of several measures of response, including visual analysis, SUV_{max} and the volumetric measures TV and TLG have been tested using pre- and post- therapy PET scans of 85 patients with DLBCL undergoing R-CHOP or R-CEOP chemotherapy. The main aim of the analysis was to see if any of these measures would prove to be beneficial for predicting response to therapy with the potential to be used to alter treatment in non-responding patients to improve their chances of survival. Visual analysis was done by a consulting physician who also aided the segmentation of disease, using a fixed 2.5 SUV threshold region growing algorithm, to obtain TV and TLG. SUV_{max} was taken as the maximum voxel within the TV. All these measures, along with Ann Arbor staging and IPI were used to assess OS and PFS in all 85 patients with DLBCL.

Statistical analysis showed some correlation between most measures and OS and PFS, however, ROC analysis suggested none of the methods were useful at predicting five-year PFS or OS. However, it may be that splitting patients into three, rather than two groups, is a more sensible approach and this is being looked into. Most patients do respond to R-CHOP or R-CEOP therapy

in NHL patients but there are patients who respond more poorly or with no prolonged response. Using three categories of responders can often divide datasets into groups in which patients have a very high response rate, a very poor response rate and a divided group between the two where it is difficult to distinguish which patients will respond and which won't. This is potentially useful as it can differentiate those patients who are likely to respond and should remain on the same treatment regime and those who are unlikely to respond should have their treatment regime changed. However, it would still leave a good amount of patients where their potential response was unknown, and therefore successfully splitting the patients into responders and non-responders, if possible, is more ideal. The initial analysis on this group of patients however failed to demonstrate that any of the clinical, visual or quantitative parameters could accurately separate patients into groups of responders and non-responders.

6) Conclusions

The aim of this project was to investigate the use of pre- and post- therapy ^{18}F-FDG PET scans for identifying response in cancer patients using both qualitative and quantitative methods. A review of the current literature has shown that ^{18}F-FDG PET can be successful at identifying response in a variety of cancers. Standardisation of imaging protocols and methods is an ever important factor in assessing response as it can reduce variation of SUVs and other quantitative parameters which can be used to predict response. PETTRA software was developed to allow users to view and analyse PET images and compute quantitative response measures, including SUV_{max}, SUV_{peak}, TV, TLG and IVH parameters. The development of this software was done with imaging protocols and methods in mind and customised for assessing quantitative measures for investigating response. It was also customised to take into account the work done on registering pre- and post- therapy PET/CT studies.

The methodology for registering pre- and post- therapy PET/CT scans, using the CT component for registration, was developed and evaluated using both qualitative and quantitative analysis. The main aims of developing this methodology were, firstly, with the view to aid consultant physicans in visual analysis and secondly, and more importantly, to potentially allow quantitative measures to be extracted from registered scans or subtraction images. Both rigid and non-rigid IRTK algorithms were customised for registering longitudinal pre- and post- therapy PET/CT scans. Registration accuracy was assessed using visual and landmark analysis on a cohort of 20 lymphoma patients. Visual analysis found IRTK non-rigid and rigid methods produced the most accurate registration methods for both PET and CT images, respectively. For the landmark analysis, misalignment errors between pre- and post- therapy images were found to be 40mm on average. Image registration was shown to substantially reduce this with average misalignments of ~10mm and ~6.5mm for IRTK rigid and non-rigid algorithms respectively. The affect these

registrations had on SUVs and TV in PET images was also investigated. While rigid registration transformation caused insignificant changes in SUV_{max} and minimal changes in SUV_{peak} and TV (<2%), non-rigid transformations caused changes in SUV_{max} (~2%) and significant changes in SUV_{peak} and TV (mean changes of 11% and 21% respectively).

The registration methodology could potentially be used to aid consultant physicans given its accuracy. It could also be used, as planned, to produce subtraction images of pre- and post-therapy PET scans which can then be quantified on a voxel-by-voxel basis to assess tumours. Although, there are many potential pitfalls with this technique there is still a benefit in at least investigating its potential. This was not investigating in the thesis but is certainly an area of future work which would be a beneficial area of research. Using the registration methodology and extracting semi-quantative parameters for assessing response on registered and subtraction images in the two datasets of mesothelioma and DLBCL patients could prove to produce more novel and accurate measures of predicting the survival of both sets of patients. While the registration was judged to be accurate, whether it is accurate enough to allow successful voxel-by-voxel analysis to produce response parameters is unknown and something that would need consideration in attempting this sort of research. Equally, even if non-rigid registration was considered to be accurate enough, the studies have shown that changes in TV would significantly influence results so this would also have to be worked around. There are clearly pitfalls to this proposed area of research and this technique, if ever to be successful, is certainly in its infancy. However, that does not mean it cannot be a valuable area of research. The PETTRA software was designed with this in mind and is set-up to display images of pre-therapy images with registered post-therapy images with the ability to segment TVs and applying them to both scans a possibility amongst other tools. Unfortunately, there was not time to do such analysis for this thesis.

Two datasets of patients with mesothelioma and DLBCL were analysed using PETTRA software to assess response measures in terms of predicting survival. Analysis of 14 patients with mesothelioma assessed a range of response parameters including SUV_{max}, SUV_{peak}, TV, TLG and IVH parameters. Different variations of these measures were also tested to try and assess the impact of different methodologies and segmentation methods on quantitative values and response prediction. SUV_{max} was computed using different normalisation methods, SUV_{peak} using different methodologies, TV and TLG using 13 different fixed threshold methods and nine adaptive threshold methods. With just 14 patients, assessment lacked statistical power, but pre- and post-therapy values of SUV_{max}, SUV_{peak}, TV and TLG showed some promise in predicting OS and PFS than the change between pre- and post- therapy values. This suggests that the volume and intensity of disease before or after therapy is more important than the change between them. IVH parameters had mixed results and while they can not be said to have showed a good indication of response, they may be worthy of further investigation. A comparison of different methodologies for obtaining SUV_{max} and SUV_{peak} parameters and different segmentation methods for obtained TV and TLG showed no real difference between them and their ability to predict OS or PFS.

Unfortunately, a lack of patients make it hard to draw any definite conclusions from the dataset of mesothelioma patients, with just 14 patients it is much more likely a parameter will work out of chance rather than because it is an accurate measure of response. An additional issue with the dataset is that the type of disease and effect of treatment is very different when compared to a disease like lymphoma where there are more responders. With the mesothelioma patients, no patients responded to treatment to reach a complete response, so parameters were merely trying to identify patients with longer survival times rather than those which had a complete response. While this is still valuable, it means that the results cannot be considered to be at all likely to correspond with results from other diseases compared to a number of studies on different types of lymphoma, for example.

The most interesting findings from the analysis on the group of mesothelioma patients was the comparison between different measures of response in terms of using different segmentation methods for TV and TLG measures and using different units and methodology for SUV_{max} and SUV_{peak} values. While once again, conclusions must be considered tentative due to a lack of patients, the results suggest that the methodology and segmentation methods used have very little influence on the success of these measures to predict survival with all showing very similar results when being statistically compared with PFS or OS. However, doing a similar study on the DLBCL dataset with more patients available would be beneficial to see if this remains the case over a larger dataset or a significantly different type. The evaluation of segmentation methods on the mesothelioma dataset was a useful comparison but perhaps the number of methods used, particularly the number of fixed threshold methods used, could have been reduced and the time used to investigate more segmentation methods on the DLBCL dataset.

Analysis of 85 patients with DLBCL measured SUV_{max}, TV and TLG on pre- and post- therapy PET scans with segmentation achieved using a fixed 2.5 SUV threshold region growing algorithm, to obtain TV and TLG. Statistical analysis showed some correlation between most measures and OS and PFS, however, ROC analysis revealed that none of the methods were useful at predicting five-year PFS or OS. Further work on this dataset is still ongoing with more response measures being applied as, unlike the mesothelioma dataset, it has a large number of patients and, therefore, more statistical power. Interestingly, along with the mesothelioma dataset, pre- and post- therapy parameters alone appear to be good indicators of patient survival, suggesting that the actual response seen from the pre-therapy to post-therapy scan is not as relevant as the volume or intensity of disease either before or after one cycle of chemotherapy.

The analysis on the 85 DLBCL patients has shown that the parameters used are not capable of simple dividing patients into groups of responders and non-responders to treatment. However, further statistical analysis is needed to investigate what the data does show. It may be possible that they can still divide patients into potentially worthwhile categories or produce statistical analysis which shows that some parameters may be more valuable than others. Unfortunately, there was not time to do this to incorporate the work into this thesis. In addition to this, further work on registration should be worked into the analysis of the two patients groups studied and used to investigate other response parameters, along with those used in the mesothelioma dataset. There is clearly more work to be done on the DLBCL data, especially as this dataset has grown and there are now over 100 patients that can be used for analysis.

References

ABD EL-HAFEZ, Y. G., MOUSTAFA, H. M., KHALIL, H. F., LIAO, C. T. & YEN, T. C. 2012. Total lesion glycolysis: A possible new prognostic parameter in oral cavity squamous cell carcinoma. *Oral Oncol*.

ADAMS, A. H., BROOKEMAN, J. R. & MERICKEL, M. B. 1991. Breast Lesion Discrimination Using Statistical-Analysis and Shape Measures on Magnetic-Resonance Imagery. *Computerized Medical Imaging and Graphics,* 15**,** 339-349.

AKHURST, T., NG, V. V., LARSON, S. M., O'DONOGHUE, J. A., O'NEEL, J., ERDI, Y. & DIVGI, C. R. 2000. Tumor Burden Assessment with Positron Emission Tomography with. *Clin Positron Imaging,* 3**,** 57-65.

ALVARENGA, A. V., PEREIRA, W. C. A., INFANTOSI, A. F. C. & AZEVEDO, C. M. 2007. Complexity curve and grey level co-occurrence matrix in the texture evaluation of breast tumor on ultrasound images. *Med Phys,* 34**,** 379-387.

AMORIM, B. J., RAMOS, C. D., DOS SANTOS, A. O., DE LIMA MDA, C., MIN, L. L., CAMARGO, E. E., CENDES, F. & ETCHEBEHERE, E. C. 2010. Brain SPECT in mesial temporal lobe epilepsy: comparison between visual analysis and SPM. *Arq Neuropsiquiatr,* 68**,** 153-60.

ANDREOPOULOU, E., ROSS, P. J., O'BRIEN, M. E. R., FORD, H. E. R., PRIEST, K., EISEN, T., NORTON, A., ASHLEY, S. & SMITH, I. E. 2004. The palliative benefits of MVP (mitomycin C, vinblastine and cisplatin) chemotherapy in patients with malignant mesothelioma. *Annals of Oncology,* 15**,** 1406-1412.

ANTOCH, G., JENTZEN, W., FREUDENBERG, L. S., STATTAUS, J., MUELLER, S. P., DEBATIN, J. F. & BOCKISCH, A. 2003. Effect of oral contrast agents on computed tomography-based positron emission tomography attenuation correction in dual-modality positron emission tomography/computed tomography imaging. *Invest Radiol,* 38**,** 784-9.

ARSLAN, N., TUNCEL, M., KUZHAN, O., ALAGOZ, E., BUDAKOGLU, B., OZET, A. & OZGUVEN, M. A. 2011. Evaluation of outcome prediction and disease extension by quantitative 2-deoxy-2-[18F] fluoro-D-glucose with positron emission tomography in patients with small cell lung cancer. *Ann Nucl Med,* 25**,** 406-13.

ASSELIN, M. C., CUNNINGHAM, V. J., AMANO, S., GUNN, R. N. & NAHMIAS, C. 2004. Parametrically defined cerebral blood vessels as non-invasive blood input functions for brain PET studies. *Phys Med Biol,* 49**,** 1033-54.

ASTON, J. A., CUNNINGHAM, V. J., ASSELIN, M. C., HAMMERS, A., EVANS, A. C. & GUNN, R. N. 2002. Positron emission tomography partial volume correction: estimation and algorithms. *J Cereb Blood Flow Metab,* 22**,** 1019-34.

AVRIL, N., BENSE, S., ZIEGLER, S. I., DOSE, J., WEBER, W., LAUBENBACHER, C., ROMER, W., JANICKE, F. & SCHWAIGER, M. 1997. Breast imaging with fluorine-18-FDG PET: Quantitative image analysis. *Journal of Nuclear Medicine,* 38**,** 1186-1191.

AVRIL, N., SASSEN, S. & ROYLANCE, R. 2009. Response to Therapy in Breast Cancer. *Journal of Nuclear Medicine,* 50**,** 55s-63s.

BACHARACH, S. L. 2007. PET/CT attenuation correction: Breathing lessons. *Journal of Nuclear Medicine,* 48**,** 677-679.

BALDINI, E. H. 2009. Radiation therapy options for malignant pleural mesothelioma. *Semin Thorac Cardiovasc Surg,* 21**,** 159-63.

BARALDI, A. & PARMIGGIANI, F. 1995. An Investigation of the Textural Characteristics Associated with Gray-Level Cooccurrence Matrix Statistical Parameters. *Ieee Transactions on Geoscience and Remote Sensing,* 33**,** 293-304.

BARRINGTON, S. F., QIAN, W. D., SOMER, E. J., FRANCESCHETTO, A., BAGNI, B., BRUN, E., ALMQUIST, H., LOFT, A., HOJGAARD, L., FEDERICO, M., GALLAMINI, A., SMITH, P., JOHNSON, P., RADFORD, J. & O'DOHERTY, M. J. 2010. Concordance between

four European centres of PET reporting criteria designed for use in multicentre trials in Hodgkin lymphoma. *Eur J Nucl Med Mol Imaging,* 37**,** 1824-1833.

BASU, S. & ALAVI, A. 2007. Partial volume correction of standardized uptake values and the dual time point in FDG-PET imaging: should these be routinely employed in assessing patients with cancer? *Eur J Nucl Med Mol Imaging,* 34**,** 1527-9.

BASU, S., SABOURY, B., TORIGIAN, D. A. & ALAVI, A. 2011. Current Evidence Base of FDG-PET/CT Imaging in the Clinical Management of Malignant Pleural Mesothelioma: Emerging Significance of Image Segmentation and Global Disease Assessment. *Molecular Imaging and Biology,* 13**,** 801-811.

BASU, S., ZAIDI, H., HOUSENI, M., BURAL, G., UDUPA, J., ACTON, P., TORIGIAN, D. A. & ALAVI, A. 2007. Novel quantitative techniques for assessing regional and global function and structure based on modern imaging modalities: implications for normal variation, aging and diseased states. *Semin Nucl Med,* 37**,** 223-39.

BENARD, F., STERMAN, D., SMITH, R. J., KAISER, L. R., ALBELDA, S. M. & ALAVI, A. 1998. Metabolic imaging of malignant pleural mesothelioma with fluorodeoxyglucose positron emission tomography. *Chest,* 114**,** 713-722.

BENARD, F., STERMAN, D., SMITH, R. J., KAISER, L. R., ALBELDA, S. M. & ALAVI, A. 1999. Prognostic value of FDG PET imaging in malignant pleural mesothelioma. *J Nucl Med,* 40**,** 1241-5.

BENCHERIF, B., STUMPF, M. J., LINKS, J. M. & FROST, J. J. 2004. Application of MRI-based partial-volume correction to the analysis of PET images of mu-opioid receptors using statistical parametric mapping. *J Nucl Med,* 45**,** 402-8.

BEN-HAIM, S. & ELL, P. 2009. (18)F-FDG PET and PET/CT in the Evaluation of Cancer Treatment Response. *Journal of Nuclear Medicine,* 50**,** 88-99.

BENZ, M. R., ALLEN-AUERBACH, M. S., EILBER, F. C., CHEN, H. J., DRY, S., PHELPS, M. E., CZERNIN, J. & WEBER, W. A. 2008. Combined assessment of metabolic and volumetric

changes for assessment of tumor response in patients with soft-tissue sarcomas. *J Nucl Med,* 49**,** 1579-84.

BERKOWITZ, A., BASU, S., SRINIVAS, S., SANKARAN, S., SCHUSTER, S. & ALAVI, A. 2008. Determination of whole-body metabolic burden as a quantitative measure of disease activity in lymphoma: a novel approach with fluorodeoxyglucose-PET. *Nucl Med Commun,* 29**,** 521-6.

BEYER, T., CZERNIN, J. & FREUDENBERG, L. S. 2011. Variations in clinical PET/CT operations: results of an international survey of active PET/CT users. *J Nucl Med,* 52**,** 303-10.

BEYER, T., TOWNSEND, D. W., BRUN, T., KINAHAN, P. E., CHARRON, M., RODDY, R., JERIN, J., YOUNG, J., BYARS, L. & NUTT, R. 2000. A combined PET/CT scanner for clinical oncology. *J Nucl Med,* 41**,** 1369-79.

BIEHL, K. J., KONG, F. M., DEHDASHTI, F., JIN, J. Y., MUTIC, S., EL NAQA, I., SIEGEL, B. A. & BRADLEY, J. D. 2006. 18F-FDG PET definition of gross tumor volume for radiotherapy of non-small cell lung cancer: is a single standardized uptake value threshold approach appropriate? *J Nucl Med,* 47**,** 1808-12.

BLACK, Q. C., GRILLS, I. S., KESTIN, L. L., WONG, C. Y. O., WONG, J. W., MARTINEZ, A. A. & YAN, D. 2004. Defining a radiotherapy target with positron emission tomography. *International Journal of Radiation Oncology Biology Physics,* 60**,** 1272-1282.

BOELLAARD, R. 2009. Standards for PET image acquisition and quantitative data analysis. *J Nucl Med,* 50 Suppl 1**,** 11S-20S.

BOELLAARD, R. 2011. Need for standardization of 18F-FDG PET/CT for treatment response assessments. *J Nucl Med,* 52 Suppl 2**,** 93S-100S.

BOELLAARD, R., KRAK, N. C., HOEKSTRA, O. S. & LAMMERTSMA, A. A. 2004. Effects of noise, image resolution, and ROI definition on the accuracy of standard uptake values: a simulation study. *J Nucl Med,* 45**,** 1519-27.

BOELLAARD, R., O'DOHERTY, M. J., WEBER, W. A., MOTTAGHY, F. M., LONSDALE, M. N., STROOBANTS, S. G., OYEN, W. J., KOTZERKE, J., HOEKSTRA, O. S., PRUIM, J., MARSDEN, P. K., TATSCH, K., HOEKSTRA, C. J., VISSER, E. P., ARENDS, B., VERZIJLBERGEN, F. J., ZIJLSTRA, J. M., COMANS, E. F., LAMMERTSMA, A. A., PAANS, A. M., WILLEMSEN, A. T., BEYER, T., BOCKISCH, A., SCHAEFER-PROKOP, C., DELBEKE, D., BAUM, R. P., CHITI, A. & KRAUSE, B. J. 2010. FDG PET and PET/CT: EANM procedure guidelines for tumour PET imaging: version 1.0. *Eur J Nucl Med Mol Imaging,* 37**,** 181-200.

BOELLAARD, R., OYEN, W. J., HOEKSTRA, C. J., HOEKSTRA, O. S., VISSER, E. P., WILLEMSEN, A. T., ARENDS, B., VERZIJLBERGEN, F. J., ZIJLSTRA, J., PAANS, A. M., COMANS, E. F. & PRUIM, J. 2008. The Netherlands protocol for standardisation and quantification of FDG whole body PET studies in multi-centre trials. *Eur J Nucl Med Mol Imaging,* 35**,** 2320-33.

BOTTOMLEY, A., GAAFA, R., MANEGOLD, C., BURGERS, S., COENS, C., LEGRAND, C., VINCENT, M., GIACCONE, G. & VAN MEERBEECK, J. 2006. Short-term treatment-related symptoms and quality of life: Results from an international randomized phase III study of cisplatin with or without raltitrexed in patients with malignant pleural mesothelioma: An EORTC Lung-Cancer Group and National Cancer Institute, Canada, Intergroup Study. *Journal of Clinical Oncology,* 24**,** 1435-1442.

BOUCEK, J. A., FRANCIS, R. J., JONES, C. G., KHAN, N., TURLACH, B. A. & GREEN, A. J. 2008. Assessment of tumour response with (18)F-fluorodeoxyglucose positron emission tomography using three-dimensional measures compared to SUVmax--a phantom study. *Phys Med Biol,* 53**,** 4213-30.

BRENNER, W., VERNON, C., MUZI, M., MANKOFF, D. A., LINK, J. M., CONRAD, E. U. & EARY, J. F. 2004. Comparison of different quantitative approaches to 18F-fluoride PET scans. *J Nucl Med,* 45**,** 1493-500.

BREPOELS, L., STROOBANTS, S. & VERHOEF, G. 2007. PET and PET/CT for response evaluation in lymphoma: Current practice and developments. *Leuk Lymphoma,* 48**,** 270-282.

BROWNELL, G. L. & SWEET, W. H. 1953. Localization of brain tumors with positron emitters. *Nucleonics,* 11**,** 40-45.

BURAL, G. G., TORIGIAN, D. A., CHAMROONRAT, W., ALKHAWALDEH, K., HOUSENI, M., EL-HADDAD, G. & ALAVI, A. 2006. Quantitative assessment of the atherosclerotic burden of the aorta by combined FDG-PET and CT image analysis: a new concept. *Nucl Med Biol,* 33**,** 1037-43.

BUSSINK, J., KAANDERS, J. H. A. M., VAN DER GRAAF, W. T. A. & OYEN, W. J. G. 2011. PET-CT for radiotherapy treatment planning and response monitoring in solid tumors. *Nature Reviews Clinical Oncology,* 8**,** 233-242.

BYRNE, M. J. & NOWAK, A. K. 2004. Modified RECIST criteria for assessment of response in malignant pleural mesothelioma. *Annals of Oncology,* 15**,** 257-260.

CARBONE, P. P., KAPLAN, H. S., MUSSHOFF, K., SMITHERS, D. W. & TUBIANA, M. 1971. Report of the Committee on Hodgkin's Disease Staging Classification. *Cancer Research,* 31**,** 1860-1.

CASASNOVAS, R. O., MEIGNAN, M., BERRIOLO-RIEDINGER, A., BARDET, S., JULIAN, A., THIEBLEMONT, C., VERA, P., BOLOGNA, S., BRIERE, J., JAIS, J. P., HAIOUN, C., COIFFIER, B. & MORSCHHAUSER, F. 2011. SUVmax reduction improves early prognosis value of interim positron emission tomography scans in diffuse large B-cell lymphoma. *Blood,* 118**,** 37-43.

CASASNOVAS, R. O., MEIGNAN, M., BERRIOLO-RIEDINGER, A., ITTI, E., HUGLO, D., HAIOUN, C. & MORSCHHAUSER, F. 2012. Early interim PET scans in diffuse large B-cell lymphoma: can there be consensus about standardized reporting, and can PET scans guide therapy choices? *Curr Hematol Malig Rep,* 7**,** 193-9.

CASEY, M. E. & NUTT, R. 1986. A multicrystal two dimensional BGO detector system for Positron Emission Tomography. *IEEE Transactions on Nuclear Science,* 33**,** 460-463.

CASHEN, A. F., DEHDASHTI, F., LUO, J. Q., HOMB, A., SIEGEL, B. A. & BARTLETT, N. L. 2011. F-18-FDG PET/CT for Early Response Assessment in Diffuse Large B-Cell Lymphoma: Poor Predictive Value of International Harmonization Project Interpretation. *Journal of Nuclear Medicine,* 52**,** 386-392.

CASTELL, F. & COOK, G. J. 2008. Quantitative techniques in 18FDG PET scanning in oncology. *Br J Cancer,* 98**,** 1597-601.

CASTELLANO, G., BONILHA, L., LI, L. M. & CENDES, F. 2004. Texture analysis of medical images. *Clin Radiol,* 59**,** 1061-1069.

CERESOLI, G. L., CHITI, A., ZUCALI, P. A., RODARI, M., LUTMAN, R. F., SALAMINA, S., INCARBONE, M., ALLOISIO, M. & SANTORO, A. 2006. Early response evaluation in malignant pleural mesothelioma by positron emission tomography with [18F]fluorodeoxyglucose. *J Clin Oncol,* 24**,** 4587-93.

CERESOLI, G. L., ZUCALI, P. A., GIANONCELLI, L., LORENZI, E. & SANTORO, A. 2010. Second-line treatment for malignant pleural mesothelioma. *Cancer Treat Rev,* 36**,** 24-32.

CHAN, W. C., ARMITAGE, J. O., GASCOYNE, R., CONNORS, J., CLOSE, P., JACOBS, P., NORTON, A., LISTER, T. A., PEDRINIS, E., CAVALLI, F., BERGER, F., COIFFIER, B., HO, F., LIANG, R., SCHAUER, A., HIDDEMANN, W., DIEBOLD, J., MACLENNAN, K. A., MULLERHERMELINK, H. K., NATHWANI, B. N., WEISENBURGER, D. D., HARRIS, N. L., ANDERSON, J. R. & ROY, P. 1997. A clinical evaluation of the International Lymphoma Study Group classification of non-Hodgkin's lymphoma. *Blood,* 89**,** 3909-3918.

CHEEBSUMON, P., VELASQUEZ, L. M., HOEKSTRA, C. J., HAYES, W., KLOET, R. W., HOETJES, N. J., SMIT, E. F., HOEKSTRA, O. S., LAMMERTSMA, A. A. & BOELLAARD, R. 2011. Measuring response to therapy using FDG PET: semi-quantitative and full kinetic analysis. *Eur J Nucl Med Mol Imaging,* 38**,** 832-42.

CHEN, H. H., CHIU, N. T., SU, W. C., GUO, H. R. & LEE, B. F. 2012. Prognostic value of whole-body total lesion glycolysis at pretreatment FDG PET/CT in non-small cell lung cancer. *Radiology,* 264**,** 559-66.

CHERRY, S. R. 2001. Fundamentals of positron emission tomography and applications in preclinical drug development. *J Clin Pharmacol,* 41**,** 482-91.

CHERRY, S. R. 2006. The 2006 Henry N. Wagner Lecture: Of mice and men (and positrons)--advances in PET imaging technology. *J Nucl Med,* 47**,** 1735-45.

CHESON, B. D., HORNING, S. J., COIFFIER, B., SHIPP, M. A., FISHER, R. I., CONNORS, J. M., LISTER, T. A., VOSE, J., GRILLO-LOPEZ, A., HAGENBEEK, A., CABANILLAS, F., KLIPPENSTEN, D., HIDDEMANN, W., CASTELLINO, R., HARRIS, N. L., ARMITAGE, J. O., CARTER, W., HOPPE, R. & CANELLOS, G. P. 1999. Report of an international workshop to standardize response criteria for non-Hodgkin's lymphomas. NCI Sponsored International Working Group. *J Clin Oncol,* 17**,** 1244.

CHESON, B. D., PFISTNER, B., JUWEID, M. E., GASCOYNE, R. D., SPECHT, L., HORNING, S. J., COIFFIER, B., FISHER, R. I., HAGENBEEK, A., ZUCCA, E., ROSEN, S. T., STROOBANTS, S., LISTER, T. A., HOPPE, R. T., DREYLING, M., TOBINAI, K., VOSE, J. M., CONNORS, J. M., FEDERICO, M. & DIEHL, V. 2007. Revised response criteria for malignant lymphoma. *J Clin Oncol,* 25**,** 579-86.

CHI, P. C., MAWLAWI, O., NEHMEH, S. A., ERDI, Y. E., BALTER, P. A., LUO, D., MOHAN, R. & PAN, T. 2007. Design of respiration averaged CT for attenuation correction of the PET data from PET/CT. *Med Phys,* 34**,** 2039-47.

CHICKLORE, S., GOH, V., SIDDIQUE, M., ROY, A., MARSDEN, P. K. & COOK, G. J. 2013. Quantifying tumour heterogeneity in (18)F-FDG PET/CT imaging by texture analysis. *Eur J Nucl Med Mol Imaging,* 40**,** 133-40.

CHIHARA, D., OKI, Y., ONODA, H., TAJI, H., YAMAMOTO, K., TAMAKI, T. & MORISHIMA, Y. 2011. High maximum standard uptake value (SUVmax) on PET scan is

associated with shorter survival in patients with diffuse large B cell lymphoma. *Int J Hematol,* 93**,** 502-8.

CHOI, H., CHARNSANGAVEJ, C., FARIA, S. C., MACAPINLAC, H. A., BURGESS, M. A., PATEL, S. R., CHEN, L. L., PODOLOFF, D. A. & BENJAMIN, R. S. 2007. Correlation of computed tomography and positron emission tomography in patients with metastatic gastrointestinal stromal tumor treated at a single institution with imatinib mesylate: proposal of new computed tomography response criteria. *J Clin Oncol,* 25**,** 1753-9.

CHUNG, H. H., KWON, H. W., KANG, K. W., PARK, N. H., SONG, Y. S., CHUNG, J. K., KANG, S. B. & KIM, J. W. 2012. Prognostic value of preoperative metabolic tumor volume and total lesion glycolysis in patients with epithelial ovarian cancer. *Ann Surg Oncol,* 19**,** 1966-72.

COHADE, C., OSMAN, M., NAKAMOTO, Y., MARSHALL, L. T., LINKS, J. M., FISHMAN, E. K. & WAHL, R. L. 2003. Initial experience with oral contrast in PET/CT: phantom and clinical studies. *J Nucl Med,* 44**,** 412-6.

COIFFIER, B., LEPAGE, E., BRIERE, J., HERBRECHT, R., TILLY, H., BOUABDALLAH, R., MOREL, P., VAN DEN NESTE, E., SALLES, G., GAULARD, P., REYES, F., LEDERLIN, P. & GISSELBRECHT, C. 2002. CHOP chemotherapy plus rituximab compared with CHOP alone in elderly patients with diffuse large-B-cell lymphoma. *N Engl J Med,* 346**,** 235-42.

COIFFIER, B., THIEBLEMONT, C., VAN DEN NESTE, E., LEPEU, G., PLANTIER, I., CASTAIGNE, S., LEFORT, S., MARIT, G., MACRO, M., SEBBAN, C., BELHADJ, K., BORDESSOULE, D., FERME, C. & TILLY, H. 2010. Long-term outcome of patients in the LNH-98.5 trial, the first randomized study comparing rituximab-CHOP to standard CHOP chemotherapy in DLBCL patients: a study by the Groupe d'Etudes des Lymphomes de l'Adulte. *Blood,* 116**,** 2040-5.

COLEMAN, R. E. 2002. Is quantitation necessary for oncological PET studies? For. *Eur J Nucl Med Mol Imaging,* 29**,** 133-5.

COSTELLOE, C. M., MACAPINLAC, H. A., MADEWELL, J. E., FITZGERALD, N. E., MAWLAWI, O. R., ROHREN, E. M., RAYMOND, A. K., LEWIS, V. O., ANDERSON, P. M., BASSETT, R. L., JR., HARRELL, R. K. & MAROM, E. M. 2009. 18F-FDG PET/CT as an indicator of progression-free and overall survival in osteosarcoma. *J Nucl Med,* 50**,** 340-7.

CRUM, W. R., GRIFFIN, L. D., HILL, D. L. G. & HAWKES, D. J. 2003. Zen and the art of medical image registration: correspondence, homology, and quality. *Neuroimage,* 20**,** 1425-1437.

DAISNE, J. F., SIBOMANA, M., BOL, A., DOUMONT, T., LONNEUX, M. & GREGOIRE, V. 2003. Tri-dimensional automatic segmentation of PET volumes based on measured source-to-background ratios: influence of reconstruction algorithms. *Radiotherapy and Oncology,* 69**,** 247-250.

DAUBE-WITHERSPOON, M. E., KARP, J. S., CASEY, M. E., DIFILIPPO, F. P., HINES, H., MUEHLLEHNER, G., SIMCIC, V., STEARNS, C. W., ADAM, L. E., KOHLMYER, S. & SOSSI, V. 2002. PET performance measurements using the NEMA NU 2-2001 standard. *J Nucl Med,* 43**,** 1398-409.

DAVIS, J. B., REINER, B., HUSER, M., BURGER, C., SZEKELY, G. & CIERNIK, I. F. 2006. Assessment of 18F PET signals for automatic target volume definition in radiotherapy treatment planning. *Radiother Oncol,* 80**,** 43-50.

DE LANGEN, A. J., VINCENT, A., VELASQUEZ, L. M., VAN TINTEREN, H., BOELLAARD, R., SHANKAR, L. K., BOERS, M., SMIT, E. F., STROOBANTS, S., WEBER, W. A. & HOEKSTRA, O. S. 2012. Repeatability of 18F-FDG uptake measurements in tumors: a metaanalysis. *J Nucl Med,* 53**,** 701-8.

DE MOOR, K., NUYTS, J., PLESSERS, L., STROOBANTS, S., MAES, F. & DUPONT, P. 2006. Non-rigid registration with position dependent rigidity for whole body PET follow-up studies. *2006 IEEE Nuclear Science Symposium Conference Record*.

DE RUYSSCHER, D., NESTLE, U., JERAJ, R. & MACMANUS, M. 2012. PET scans in radiotherapy planning of lung cancer. *Lung Cancer,* 75**,** 141-5.

DEGRADO, T. R., COLEMAN, R. E., WANG, S., BALDWIN, S. W., ORR, M. D., ROBERTSON, C. N., POLASCIK, T. J. & PRICE, D. T. 2001. Synthesis and evaluation of 18F-labeled choline as an oncologic tracer for positron emission tomography: initial findings in prostate cancer. *Cancer Research,* 61**,** 110-7.

DELBEKE, D., COLEMAN, R. E., GUIBERTEAU, M. J., BROWN, M. L., ROYAL, H. D., SIEGEL, B. A., TOWNSEND, D. W., BERLAND, L. L., PARKER, J. A., HUBNER, K., STABIN, M. G., ZUBAL, G., KACHELRIESS, M., CRONIN, V. & HOLBROOK, S. 2006. Procedure guideline for tumor imaging with 18F-FDG PET/CT 1.0. *J Nucl Med,* 47**,** 885-95.

DEWALLE-VIGNION, A. S., BETROUNI, N., LOPES, R., HUGLO, D., STUTE, S. & VERMANDEL, M. 2011. A New Method for Volume Segmentation of PET Images, Based on Possibility Theory. *IEEE Trans Med Imaging,* 30**,** 409-423.

DI CHIRO, G. & BROOKS, R. A. 1988. PET quantitation: blessing and curse. *J Nucl Med,* 29**,** 1603-4.

DIBBLE, E. H., ALVAREZ, A. C., TRUONG, M. T., MERCIER, G., COOK, E. F. & SUBRAMANIAM, R. M. 2012. 18F-FDG metabolic tumor volume and total glycolytic activity of oral cavity and oropharyngeal squamous cell cancer: adding value to clinical staging. *J Nucl Med,* 53**,** 709-15.

DIEDERICHS, C. G., STAIB, L., GLATTING, G., BEGER, H. G. & RESKE, S. N. 1998. FDG PET: elevated plasma glucose reduces both uptake and detection rate of pancreatic malignancies. *J Nucl Med,* 39**,** 1030-3.

DIMITRAKOPOULOU-STRAUSS, A., STRAUSS, L. G., HEICHEL, T., WU, H., BURGER, C., BERND, L. & EWERBECK, V. 2002. The role of quantitative (18)F-FDG PET studies for the differentiation of malignant and benign bone lesions. *J Nucl Med,* 43**,** 510-8.

DIZENDORF, E., HANY, T. F., BUCK, A., VON SCHULTHESS, G. K. & BURGER, C. 2003. Cause and magnitude of the error induced by oral CT contrast agent in CT-based attenuation correction of PET emission studies. *J Nucl Med,* 44**,** 732-8.

DOOT, R. K., SCHEUERMANN, J. S., CHRISTIAN, P. E., KARP, J. S. & KINAHAN, P. E. 2010. Instrumentation factors affecting variance and bias of quantifying tracer uptake with PET/CT. *Med Phys,* 37**,** 6035-46.

DREVER, L., ROBINSON, D. M., MCEWAN, A. & ROA, W. 2006. A local contrast based approach to threshold segmentation for PET target volume delineation. *Med Phys,* 33**,** 1583-94.

DUNPHY, M. P. S. & LEWIS, J. S. 2009. Radiopharmaceuticals in preclinical and clinical development for monitoring of therapy with PET. *J Nucl Med,* 50 Suppl 1**,** 106S-21S.

DUPAS, B., AUGEUL-MEUNIER, K., FRAMPAS, E., BODET-MILIN, C., GASTINNE, T. & LE GOUILL, S. 2013. Staging and monitoring in the treatment of lymphomas. *Diagn Interv Imaging.*

DUPUIS, J., ITTI, E., RAHMOUNI, A., HEMERY, F., GISSELBRECHT, C., LIN, C., COPIE-BERGMAN, C., BELHADJ, K., EL GNAOUI, T., GAILLARD, I., KUHNOWSKI, F., MEIGNAN, M. & HAIOUN, C. 2009. Response assessment after an inductive CHOP or CHOP-like regimen with or without rituximab in 103 patients with diffuse large B-cell lymphoma: integrating 18fluorodeoxyglucose positron emission tomography to the International Workshop Criteria. *Ann Oncol,* 20**,** 503-7.

DUXBURY, D. M., OTT, R. J., FLOWER, M. A., ERLANDSSON, K., READER, A. J., BATEMAN, J. E., STEPHENSON, R. & SPILL, E. J. 1999. Preliminary results from the new large-area PETRRA positron camera. *IEEE Transactions on Nuclear Science,* 46**,** 1050-1054.

EARY, J. F., O'SULLIVAN, F., O'SULLIVAN, J. & CONRAD, E. U. 2008. Spatial heterogeneity in sarcoma 18F-FDG uptake as a predictor of patient outcome. *J Nucl Med,* 49**,** 1973-9.

EISENHAUER, E. A., THERASSE, P., BOGAERTS, J., SCHWARTZ, L. H., SARGENT, D., FORD, R., DANCEY, J., ARBUCK, S., GWYTHER, S., MOONEY, M., RUBINSTEIN, L., SHANKAR, L., DODD, L., KAPLAN, R., LACOMBE, D. & VERWEIJ, J. 2009. New response

evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer,* 45**,** 228-47.

EL FAKHRI, G., SURTI, S., TROTT, C. M., SCHEUERMANN, J. & KARP, J. S. 2011. Improvement in Lesion Detection with Whole-Body Oncologic Time-of-Flight PET. *Journal of Nuclear Medicine,* 52**,** 347-353.

EL NAQA, I., GRIGSBY, P., APTE, A., KIDD, E., DONNELLY, E., KHULLAR, D., CHAUDHARI, S., YANG, D., SCHMITT, M., LAFOREST, R., THORSTAD, W. & DEASY, J. O. 2009a. Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. *Pattern Recognit,* 42**,** 1162-1171.

EL NAQA, I., VAIDYA, M., APTE, A., DEHDASHTI, F., DEASY, J. O. & BRADLEY, J. D. 2009b. Predicting Response in Lung Cancer from FDG-PET Uptake Characteristics. *IFMBE Proceedings,* 25**,** 656–659.

EL NAQA, I., YANG, D., APTE, A., KHULLAR, D., MUTIC, S., ZHENG, J., BRADLEY, J. D., GRIGSBY, P. & DEASY, J. O. 2007. Concurrent multimodality image segmentation by active contours for radiotherapy treatment planning. *Med Phys,* 34**,** 4738-4749.

ELSTROM, R., GUAN, L., BAKER, G., NAKHODA, K., VERGILIO, J. A., ZHUANG, H., PITSILOS, S., BAGG, A., DOWNS, L., MEHROTRA, A., KIM, S., ALAVI, A. & SCHUSTER, S. J. 2003. Utility of FDG-PET scanning in lymphoma by WHO classification. *Blood,* 101**,** 3875-6.

ERASMUS, J. J., GLADISH, G. W., BROEMELING, L., SABLOFF, B. S., TRUONG, M. T., HERBST, R. S. & MUNDEN, R. F. 2003. Interobserver and intraobserver variability in measurement of non-small-cell carcinoma lung lesions: implications for assessment of tumor response. *J Clin Oncol,* 21**,** 2574-82.

ERDI, Y. E., MAWLAWI, O., LARSON, S. M., IMBRIACO, M., YEUNG, H., FINN, R. & HUMM, J. L. 1997. Segmentation of lung lesion volume by adaptive positron emission tomography image thresholding. *Cancer,* 80**,** 2505-2509.

ERDI, Y. E., NEHMEH, S. A., PAN, T., PEVSNER, A., ROSENZWEIG, K. E., MAGERAS, G., YORKE, E. D., SCHODER, H., HSIAO, W., SQUIRE, O. D., VERNON, P., ASHMAN, J. B., MOSTAFAVI, H., LARSON, S. M. & HUMM, J. L. 2004. The CT motion quantitation of lung lesions and its impact on PET-measured SUVs. *J Nucl Med,* 45**,** 1287-92.

ERSELCAN, T., TURGUT, B., DOGAN, D. & OZDEMIR, S. 2002. Lean body mass-based standardized uptake value, derived from a predictive equation, might be misleading in PET studies. *Eur J Nucl Med Mol Imaging,* 29**,** 1630-8.

EVERAERT, H., HOORENS, A., VANHOVE, C., SERMEUS, A., CEULEMANS, G., ENGELS, B., VERMEERSCH, M., VERELLEN, D., URBAIN, D., STORME, G. & DE RIDDER, M. 2011. Prediction of Response to Neoadjuvant Radiotherapy in Patients with Locally Advanced Rectal Cancer by Means of Sequential 18fdg-Pet. *International Journal of Radiation Oncology Biology Physics,* 80**,** 91-96.

FAHEY, F. H., KINAHAN, P. E., DOOT, R. K., KOCAK, M., THURSTON, H. & POUSSAINT, T. Y. 2010. Variability in PET quantitation within a multicenter consortium. *Med Phys,* 37**,** 3660-6.

FEUGIER, P., VAN HOOF, A., SEBBAN, C., SOLAL-CELIGNY, P., BOUABDALLAH, R., FERME, C., CHRISTIAN, B., LEPAGE, E., TILLY, H., MORSCHHAUSER, F., GAULARD, P., SALLES, G., BOSLY, A., GISSELBRECHT, C., REYES, F. & COIFFIER, B. 2005. Long-term results of the R-CHOP study in the treatment of elderly patients with diffuse large B-cell lymphoma: a study by the Groupe d'Etude des Lymphomes de l'Adulte. *J Clin Oncol,* 23**,** 4117-26.

FISHER, S. G. & FISHER, R. I. 2004. The epidemiology of non-Hodgkin's lymphoma. *Oncogene,* 23**,** 6524-34.

FITZPATRICK, J. M., HILL, D. L., SHYR, Y., WEST, J., STUDHOLME, C. & MAURER, C. R., JR. 1998. Visual assessment of the accuracy of retrospective registration of MR and CT images of the brain. *IEEE Trans Med Imaging,* 17**,** 571-85.

FITZPATRICK, J. M. & WEST, J. B. 2001. The distribution of target registration error in rigid-body point-based registration. *IEEE Trans Med Imaging,* 20**,** 917-27.

FITZPATRICK, J. M., WEST, J. B. & MAURER, C. R., JR. 1998. Predicting error in rigid-body point-based registration. *IEEE Trans Med Imaging,* 17**,** 694-702.

FLEISS, J. L. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin,* 76**,** 378–382.

FLETCHER, J. W., DJULBEGOVIC, B., SOARES, H. P., SIEGEL, B. A., LOWE, V. J., LYMAN, G. H., COLEMAN, R. E., WAHL, R., PASCHOLD, J. C., AVRILL, N., EINHORN, L. H., SUH, W. W., SAMSON'O, D., DELBEKELL, D., GORMAN, M. & SHIELDS, A. F. 2008. Recommendations on the use of F-18-FDG PET in oncology. *Journal of Nuclear Medicine,* 49**,** 480-508.

FLORES, R. M. 2005. The role of PET in the surgical management of malignant pleural mesothelioma. *Lung Cancer,* 49 Suppl 1**,** S27-32.

FRANCIS, R. J., BYRNE, M. J., VAN DER SCHAAF, A. A., BOUCEK, J. A., NOWAK, A. K., PHILLIPS, M., PRICE, R., PATRIKEOS, A. P., MUSK, A. W. & MILLWARD, M. J. 2007. Early prediction of response to chemotherapy and survival in malignant pleural mesothelioma using a novel semiautomated 3-dimensional volume-based analysis of serial 18F-FDG PET scans. *J Nucl Med,* 48**,** 1449-58.

FREEDMAN, N. M., SUNDARAM, S. K., KURDZIEL, K., CARRASQUILLO, J. A., WHATLEY, M., CARSON, J. M., SELLERS, D., LIBUTTI, S. K., YANG, J. C. & BACHARACH, S. L. 2003. Comparison of SUV and Patlak slope for monitoring of cancer therapy using serial PET scans. *Eur J Nucl Med Mol Imaging,* 30**,** 46-53.

FRIEDBERG, J. W., COHEN, P., CHEN, L., ROBINSON, K. S., FORERO-TORRES, A., LA CASCE, A. S., FAYAD, L. E., BESSUDO, A., CAMACHO, E. S., WILLIAMS, M. E., VAN DER JAGT, R. H., OLIVER, J. W. & CHESON, B. D. 2008. Bendamustine in patients with

rituximab-refractory indolent and transformed non-Hodgkin's lymphoma: results from a phase II multicenter, single-agent study. *J Clin Oncol,* 26**,** 204-10.

FROUIN, V., COMTAT, C., REILHAC, A. & GREGOIRE, M. C. 2002. Correction of partial-volume effect for PET striatal imaging: fast implementation and study of robustness. *J Nucl Med,* 43**,** 1715-26.

GALAVIS, P. E., HOLLENSEN, C., JALLOW, N., PALIWAL, B. & JERAJ, R. 2010. Variability of textural features in FDG PET images due to different acquisition modes and reconstruction parameters. *Acta Oncologica,* 49**,** 1012-1016.

GALLOWAY, M. M. 1975. Texture Analysis Using Gray Level Run Lengths. *Computer Graphics and Image Processing,* 4**,** 172-179.

GANESHAN, B., PANAYIOTOU, E., BURNAND, K., DIZDAREVIC, S. & MILES, K. 2012. Tumour heterogeneity in non-small cell lung carcinoma assessed by CT texture analysis: a potential marker of survival. *Eur Radiol,* 22**,** 796-802.

GANESHAN, B., SKOGEN, K., PRESSNEY, I., COUTROUBIS, D. & MILES, K. 2012. Tumour heterogeneity in oesophageal cancer assessed by CT texture analysis: Preliminary evidence of an association with tumour metabolism, stage, and survival. *Clin Radiol,* 67**,** 157-164.

GARRA, B. S., KRASNER, B. H., HORII, S. C., ASCHER, S., MUN, S. K. & ZEMAN, R. K. 1993. Improving the Distinction between Benign and Malignant Breast-Lesions - the Value of Sonographic Texture Analysis. *Ultrasonic Imaging,* 15**,** 267-285.

GAZDAR, A. F. & CARBONE, M. 2003. Molecular pathogenesis of malignant mesothelioma and its relationship to simian virus 40. *Clin Lung Cancer,* 5**,** 177-81.

GEETS, X., LEE, J. A., BOL, A., LONNEUX, M. & GREGOIRE, V. 2007. A gradient-based method for segmenting FDG-PET images: methodology and validation. *Eur J Nucl Med Mol Imaging,* 34**,** 1427-1438.

GERBAUDO, V. H., MAMEDE, M., TROTMAN-DICKENSON, B., HATABU, H. & SUGARBAKER, D. J. 2011. FDG PET/CT patterns of treatment failure of malignant pleural mesothelioma: relationship to histologic type, treatment algorithm, and survival. *Eur J Nucl Med Mol Imaging,* 38**,** 810-21.

GEWORSKI, L., KNOOP, B. O., DE WIT, M., IVANCEVIC, V., BARES, R. & MUNZ, D. L. 2002. Multicenter comparison of calibration and cross calibration of PET scanners. *J Nucl Med,* 43**,** 635-9.

GOERRES, G. W., BURGER, C., SCHWITTER, M. R., HEIDELBERG, T. N., SEIFERT, B. & VON SCHULTHESS, G. K. 2003. PET/CT of the abdomen: optimizing the patient breathing pattern. *Eur Radiol,* 13**,** 734-9.

GOERRES, G. W., HANY, T. F., KAMEL, E., VON SCHULTHESS, G. K. & BUCK, A. 2002b. Head and neck imaging with PET and PET/CT: artefacts from dental metallic implants. *Eur J Nucl Med Mol Imaging,* 29**,** 367-70.

GOERRES, G. W., KAMEL, E., HEIDELBERG, T. N., SCHWITTER, M. R., BURGER, C. & VON SCHULTHESS, G. K. 2002. PET-CT image co-registration in the thorax: influence of respiration. *Eur J Nucl Med Mol Imaging,* 29**,** 351-60.

GOERRES, G. W., ZIEGLER, S. I., BURGER, C., BERTHOLD, T., VON SCHULTHESS, G. K. & BUCK, A. 2003b. Artifacts at PET and PET/CT caused by metallic hip prosthetic material. *Radiology,* 226**,** 577-84.

GORENBERG, M., HALLETT, W. A. & O'DOHERTY, M. J. 2002. Does diabetes affect [(18)F]FDG standardised uptake values in lung cancer? *Eur J Nucl Med Mol Imaging,* 29**,** 1324-7.

GRAHAM, M. M. 2002. Is quantitation necessary for oncological PET studies? Against. *Eur J Nucl Med Mol Imaging,* 29**,** 135-8.

GRAHAM, M. M., BADAWI, R. D. & WAHL, R. L. 2011. Variations in PET/CT Methodology for Oncologic Imaging at US Academic Medical Centers: An Imaging Response Assessment Team Survey. *Journal of Nuclear Medicine,* 52**,** 311-317.

GRAHAM, M. M., PETERSON, L. M. & HAYWARD, R. M. 2000. Comparison of simplified quantitative analyses of FDG uptake. *Nucl Med Biol,* 27**,** 647-55.

GREEN, A. J., FRANCIS, R. J., BAIG, S. & BEGENT, R. H. J. 2008. Semiautomatic volume of interest drawing for F-18-FDG image analysis - method and preliminary results. *Eur J Nucl Med Mol Imaging,* 35**,** 393-406.

GREEN, S. & WEISS, G. R. 1992. Southwest Oncology Group standard response criteria, endpoint definitions and toxicity criteria. *Invest New Drugs,* 10**,** 239-53.

GRGIC, A., BALLEK, E., FLECKENSTEIN, J., MOCA, N., KREMP, S., SCHAEFER, A., KUHNIGK, J. M., RUBE, C., KIRSCH, C. M. & HELLWIG, D. 2011. Impact of rigid and nonrigid registration on the determination of 18F-FDG PET-based tumour volume and standardized uptake value in patients with lung cancer. *Eur J Nucl Med Mol Imaging,* 38**,** 856-64.

GRGIC, A., NESTLE, U., SCHAEFER-SCHULER, A., KREMP, S., BALLEK, E., FLECKENSTEIN, J., RUBE, C., KIRSCH, C. M. & HELLWIG, D. 2009. Nonrigid versus rigid registration of thoracic 18F-FDG PET and CT in patients with lung cancer: an intraindividual comparison of different breathing maneuvers. *J Nucl Med,* 50**,** 1921-6.

GULEC, S. A., SUTHAR, R. R., BAROT, T. C. & PENNINGTON, K. 2011. The prognostic value of functional tumor volume and total lesion glycolysis in patients with colorectal cancer liver metastases undergoing 90Y selective internal radiation therapy plus chemotherapy. *Eur J Nucl Med Mol Imaging,* 38**,** 1289-95.

HAIOUN, C., ITTI, E., RAHMOUNI, A., BRICE, P., RAIN, J. D., BELHADJ, K., GAULARD, P., GARDERET, L., LEPAGE, E., REYES, F. & MEIGNAN, M. 2005. [18F]fluoro-2-deoxy-D-glucose positron emission tomography (FDG-PET) in aggressive lymphoma: an early prognostic tool for predicting patient outcome. *Blood,* 106**,** 1376-81.

HALLETT, W. A. 2004. Quantification in clinical fluorodeoxyglucose positron emission tomography. *Nucl Med Commun,* 25**,** 647-50.

HALLETT, W. A., MARSDEN, P. K., CRONIN, B. F. & O'DOHERTY, M. J. 2001. Effect of corrections for blood glucose and body size on [18F]FDG PET standardised uptake values in lung cancer. *Eur J Nucl Med,* 28**,** 919-22.

HAMMOND, L. A., DENIS, L., SALMAN, U., JERABEK, P., THOMAS, C. R. & KUHN, J. G. 2003. Positron emission tomography (PET): Expanding the horizons of oncology drug development. *Investigational New Drugs,* 21**,** 309-340.

HARALICK, R. M., SHANMUGA.K & DINSTEIN, I. 1973. Textural Features for Image Classification. *Ieee Transactions on Systems Man and Cybernetics,* Smc3**,** 610-621.

HARRIS, N. L., JAFFE, E. S., DIEBOLD, J., FLANDRIN, G., MULLER-HERMELINK, H. K., VARDIMAN, J., LISTER, T. A. & BLOOMFIELD, C. D. 2000. The World Health Organization classification of neoplastic diseases of the haematopoietic and lymphoid tissues: Report of the Clinical Advisory Committee Meeting, Airlie House, Virginia, November 1997. *Histopathology,* 36**,** 69-86.

HARRIS, N. L., JAFFE, E. S., STEIN, H., BANKS, P. M., CHAN, J. K., CLEARY, M. L., DELSOL, G., DE WOLF-PEETERS, C., FALINI, B. & GATTER, K. C. 1994. A revised European-American classification of lymphoid neoplasms: a proposal from the International Lymphoma Study Group. *Blood,* 84**,** 1361-92.

HATT, M., LE REST, C. C., DESCOURT, P., DEKKER, A., DE RUYSSCHER, D., OELLERS, M., LAMBIN, P., PRADIER, O. & VISVIKIS, D. 2010. Accurate Automatic Delineation of Heterogeneous Functional Volumes in Positron Emission Tomography for Oncology Applications. *International Journal of Radiation Oncology Biology Physics,* 77**,** 301-308.

HATT, M., LE REST, C. C., TURZO, A., ROUX, C. & VISVIKIS, D. 2009. A Fuzzy Locally Adaptive Bayesian Segmentation Approach for Volume Determination in PET. *IEEE Trans Med Imaging,* 28**,** 881-893.

HATT, M., VISVIKIS, D., ALBARGHACH, N. M., TIXIER, F., PRADIER, O. & CHEZE-LE REST, C. 2011. Prognostic value of 18F-FDG PET image-based parameters in oesophageal cancer and impact of tumour delineation methodology. *Eur J Nucl Med Mol Imaging,* 38**,** 1191-202.

HAWKES, D. J. 1998. Algorithms for radiological image registration and their clinical application. *J Anat,* 193 (Pt 3)**,** 347-61.

HAZARIKA, M., WHITE, R. M., BOOTH, B. P., WANG, Y. C., HAM, D. Y. L., LIANG, C. Y., RAHMAN, A., GOBBURU, J. V. S., LI, N., SRIDHARA, R., MORSE, D. E., LOSTRITTO, R., GARVEY, P., JOHNSON, J. R. & PAZDUR, R. 2005. Pemetrexed in malignant pleural mesothelioma. *Clinical Cancer Research,* 11**,** 982-992.

HERNDON, J. E., GREEN, M. R., CHAHINIAN, A. P., CORSON, J. M., SUZUKI, Y. & VOGELZANG, N. J. 1998. Factors predictive of survival among 337 patients with mesothelioma treated between 1984 and 1994 by the Cancer and Leukemia Group B. *Chest,* 113**,** 723-731.

HICKESON, M., YUN, M., MATTHIES, A., ZHUANG, H., ADAM, L. E., LACORTE, L. & ALAVI, A. 2002. Use of a corrected standardized uptake value based on the lesion size on CT permits accurate characterization of lung nodules on FDG-PET. *Eur J Nucl Med Mol Imaging,* 29**,** 1639-47.

HICKS, R. J. 2005. The role of PET in monitoring therapy. *Cancer Imaging,* 5**,** 51-7.

HILL, D. L., BATCHELOR, P. G., HOLDEN, M. & HAWKES, D. J. 2001. Medical image registration. *Phys Med Biol,* 46**,** R1-45.

HILL, D. L., HAWKES, D. J., HUSSAIN, Z., GREEN, S. E., RUFF, C. F. & ROBINSON, G. P. 1993. Accurate combination of CT and MR data of the head: validation and applications in surgical and therapy planning. *Comput Med Imaging Graph,* 17**,** 357-63.

HODGKIN, T. 1832. On some Morbid Appearances of the Absorbent Glands and Spleen. *Med Chir Trans,* 17**,** 68-114.

HODGSON, J. T., MCELVENNY, D. M., DARNTON, A. J., PRICE, M. J. & PETO, J. 2005. The expected burden of mesothelioma mortality in Great Britain from 2002 to 2050. *Br J Cancer,* 92**,** 587-593.

HOEKSTRA, C. J., HOEKSTRA, O. S., STROOBANTS, S. G., VANSTEENKISTE, J., NUYTS, J., SMIT, E. F., BOERS, M., TWISK, J. W. & LAMMERTSMA, A. A. 2002. Methods to monitor response to chemotherapy in non-small cell lung cancer with 18F-FDG PET. *J Nucl Med,* 43**,** 1304-9.

HOFMAN, M. S. & HICKS, R. J. 2010. Restaging: Should We PERCIST Without Pattern Recognition? *Journal of Nuclear Medicine,* 51**,** 1830-1832.

HORNING, S. J., JUWEID, M. E., SCHODER, H., WISEMAN, G., MCMILLAN, A., SWINNEN, L. J., ADVANI, R., GASCOYNE, R. & QUON, A. 2010. Interim positron emission tomography scans in diffuse large B-cell lymphoma: an independent expert nuclear medicine evaluation of the Eastern Cooperative Oncology Group E3404 study. *Blood,* 115**,** 775-7; quiz 918.

HSIEH, T. C., LIN, W. Y., DING, H. J., SUN, S. S., WU, Y. C., YEN, K. Y. & KAO, C. H. 2012. Sex- and age-related differences in brain FDG metabolism of healthy adults: an SPM analysis. *J Neuroimaging,* 22**,** 21-7.

HUDSON, H. M. & LARKIN, R. S. 1994. Accelerated image reconstruction using ordered subsets of projection data. *IEEE Trans Med Imaging,* 13**,** 601-9.

HUNTER, G. J., HAMBERG, L. M., ALPERT, N. M., CHOI, N. C. & FISCHMAN, A. J. 1996. Simplified measurement of deoxyglucose utilization rate. *J Nucl Med,* 37**,** 950-5.

HUSTINX, R., SMITH, R. J., BENARD, F., ROSENTHAL, D. I., MACHTAY, M., FARBER, L. A. & ALAVI, A. 1999. Dual time point fluorine-18 fluorodeoxyglucose positron emission tomography: a potential method to differentiate malignancy from inflammation and normal tissue in the head and neck. *Eur J Nucl Med,* 26**,** 1345-8.

HUTCHINGS, M. & BARRINGTON, S. F. 2009. PET/CT for Therapy Response Assessment in Lymphoma. *Journal of Nuclear Medicine,* 50**,** 21s-30s.

HUTTON, B. F. 2011. Recent advances in iterative reconstruction for clinical SPECT/PET and CT. *Acta Oncologica,* 50**,** 851-858.

HUTTON, B. F. & BRAUN, M. 2003. Software for image registration: algorithms, accuracy, efficacy. *Semin Nucl Med,* 33**,** 180-92.

HUTTON, B. F., BRAUN, M., THURFJELL, L. & LAU, D. Y. 2002. Image registration: an essential tool for nuclear medicine. *Eur J Nucl Med Mol Imaging,* 29**,** 559-77.

HYUN, S. H., CHOI, J. Y., SHIM, Y. M., KIM, K., LEE, S. J., CHO, Y. S., LEE, J. Y., LEE, K. H. & KIM, B. T. 2010. Prognostic value of metabolic tumor volume measured by 18F-fluorodeoxyglucose positron emission tomography in patients with esophageal carcinoma. *Ann Surg Oncol,* 17**,** 115-22.

IM, H. J., KIM, T. S., PARK, S. Y., MIN, H. S., KIM, J. H., KANG, H. G., PARK, S. E., KWON, M. M., YOON, J. H., PARK, H. J., KIM, S. K. & PARK, B. K. 2012. Prediction of tumour necrosis fractions using metabolic and volumetric 18F-FDG PET/CT indices, after one course and at the completion of neoadjuvant chemotherapy, in children and young adults with osteosarcoma. *Eur J Nucl Med Mol Imaging,* 39**,** 39-49.

INTERNATIONAL NHL PROGNOSTIC FACTORS PROJECT. 1993. A predictive model for aggressive non-Hodgkin's lymphoma. *N Engl J Med,* 329**,** 987-94.

IRELAND, R. H., DYKER, K. E., BARBER, D. C., WOOD, S. M., HANNEY, M. B., TINDALE, W. B., WOODHOUSE, N., HOGGARD, N., CONWAY, J. & ROBINSON, M. H. 2007. Nonrigid image registration for head and neck cancer radiotherapy treatment planning with PET/CT. *Int J Radiat Oncol Biol Phys,* 68**,** 952-7.

ITTI, E., LIN, C., DUPUIS, J., PAONE, G., CAPACCHIONE, D., RAHMOUNI, A., HAIOUN, C. & MEIGNAN, M. 2009. Prognostic Value of Interim (18)F-FDG PET in Patients with Diffuse

Large B-Cell Lymphoma: SUV-Based Assessment at 4 Cycles of Chemotherapy. *Journal of Nuclear Medicine,* 50**,** 527-533.

JAMES, K., EISENHAUER, E., CHRISTIAN, M., TERENZIANI, M., VENA, D., MULDAL, A. & THERASSE, P. 1999. Measuring response in solid tumors: unidimensional versus bidimensional measurement. *J Natl Cancer Inst,* 91**,** 523-8.

JANNE, P. A., WANG, X. F., KRUG, L. M., HODGSON, L., VOKES, E. E. & KINDLER, H. L. 2006. Phase II trial of sorafenib (BAY 43-9006) in malignant mesothelioma: CALGB 30307. *Lung Cancer,* 54**,** S51-S52.

JASKOWIAK, C. J., BIANCO, J. A., PERLMAN, S. B. & FINE, J. P. 2005. Influence of reconstruction iterations on 18F-FDG PET/CT standardized uptake values. *J Nucl Med,* 46**,** 424-8.

JENTZEN, W., FREUDENBERG, L., EISING, E. G., HEINZE, M., BRANDAU, W. & BOCKISCH, A. 2007. Segmentation of PET volumes by iterative image thresholding. *Journal of Nuclear Medicine,* 48**,** 108-114.

JEONG, Y., CHO, S. S., PARK, J. M., KANG, S. J., LEE, J. S., KANG, E., NA, D. L. & KIM, S. E. 2005. 18F-FDG PET findings in frontotemporal dementia: an SPM analysis of 29 patients. *J Nucl Med,* 46**,** 233-9.

JERUSALEM, G., BEGUIN, Y., FASSOTTE, M. F., NAJJAR, F., PAULUS, P., RIGO, P. & FILLET, G. 1999. Whole-body positron emission tomography using F-18-fluorodeoxyglucose for posttreatment evaluation in Hodgkin's disease and non-Hodgkin's lymphoma has higher diagnostic and prognostic value than classical computed tomography scan imaging. *Blood,* 94**,** 429-433.

JERUSALEM, G., HUSTINX, R., BEGUIN, Y. & FILLET, G. 2003. PET scan imaging in oncology. *European Journal of Cancer,* 39**,** 1525-1534.

JULKA, P. K., DOVAL, D. C., GUPTA, S. & RATH, G. K. 2008. Response assessment in solid tumours: a comparison of WHO, SWOG and RECIST guidelines. *Br J Radiol,* 81**,** 444-9.

JUWEID, M. E. & CHESON, B. D. 2006. Positron-emission tomography and assessment of cancer therapy. *N Engl J Med,* 354**,** 496-507.

JUWEID, M. E., WISEMAN, G. A., VOSE, J. M., RITCHIE, J. M., MENDA, Y., WOOLDRIDGE, J. E., MOTTAGHY, F. M., ROHREN, E. M., BLUMSTEIN, N. M., STOLPEN, A., LINK, B. K., RESKE, S. N., GRAHAM, M. M. & CHESON, B. D. 2005. Response assessment of aggressive non-Hodgkin's lymphoma by integrated International Workshop Criteria and fluorine-18-fluorodeoxyglucose positron emission tomography. *J Clin Oncol,* 23**,** 4652-61.

KADRMAS, D. J., CASEY, M. E., CONTI, M., JAKOBY, B. W., LOIS, C. & TOWNSEND, D. W. 2009. Impact of time-of-flight on PET tumor detection. *J Nucl Med,* 50**,** 1315-23.

KAMEL, E. M., BURGER, C., BUCK, A., VON SCHULTHESS, G. K. & GOERRES, G. W. 2003. Impact of metallic dental implants on CT-based attenuation correction in a combined PET/CT scanner. *Eur Radiol,* 13**,** 724-8.

KASAMON, Y. L., WAHL, R. L., ZIESSMAN, H. A., BLACKFORD, A. L., GOODMAN, S. N., FIDYK, C. A., ROGERS, K. M., BOLANOS-MEADE, J., BOROWITZ, M. J., AMBINDER, R. F., JONES, R. J. & SWINNEN, L. J. 2009. Phase II study of risk-adapted therapy of newly diagnosed, aggressive non-Hodgkin lymphoma based on midtreatment FDG-PET scanning. *Biol Blood Marrow Transplant,* 15**,** 242-8.

KELLOFF, G. J., HOFFMAN, J. M., JOHNSON, B., SCHER, H. I., SIEGEL, B. A., CHENG, E. Y., CHESON, B. D., O'SHAUGHNESSY, J., GUYTON, K. Z., MANKOFF, D. A., SHANKAR, L., LARSON, S. M., SIGMAN, C. C., SCHILSKY, R. L. & SULLIVAN, D. C. 2005. Progress and promise of FDG-PET imaging for cancer patient management and oncologic drug development. *Clin Cancer Res,* 11**,** 2785-808.

KIM, C. K., GUPTA, N. C., CHANDRAMOULI, B. & ALAVI, A. 1994. Standardized uptake values of FDG: body surface area correction is preferable to body weight correction. *J Nucl Med,* 35**,** 164-7.

KIM, T. M., PAENG, J. C., CHUN, I. K., KEAM, B., JEON, Y. K., LEE, S. H., KIM, D. W., LEE, D. S., KIM, C. W., CHUNG, J. K., KIM, I. H. & HEO, D. S. 2012. Total lesion glycolysis in positron emission tomography is a better predictor of outcome than the International Prognostic Index for patients with diffuse large B cell lymphoma. *Cancer*.

KINAHAN, P. E. & FLETCHER, J. W. 2010. Positron Emission Tomography-Computed Tomography Standardized Uptake Values in Clinical Practice and Assessing Response to Therapy. *Seminars in Ultrasound Ct and Mri,* 31**,** 496-505.

KINAHAN, P. E., TOWNSEND, D. W., BEYER, T. & SASHIN, D. 1998. Attenuation correction for a combined 3D PET/CT scanner. *Med Phys,* 25**,** 2046-2053.

KLEIN, A., ANDERSSON, J., ARDEKANI, B. A., ASHBURNER, J., AVANTS, B., CHIANG, M. C., CHRISTENSEN, G. E., COLLINS, D. L., GEE, J., HELLIER, P., SONG, J. H., JENKINSON, M., LEPAGE, C., RUECKERT, D., THOMPSON, P., VERCAUTEREN, T., WOODS, R. P., MANN, J. J. & PARSEY, R. V. 2009. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuroimage,* 46**,** 786-802.

KNUUTI, J. 2008. Should we use more PET-CT in clinical cardiology? *Eur J Nucl Med Mol Imaging,* 35**,** 887-888.

KOSTAKOGLU, L., COLEMAN, M., LEONARD, J. P., KUJI, I., ZOE, H. & GOLDSMITH, S. J. 2002. PET predicts prognosis after 1 cycle of chemotherapy in aggressive lymphoma and Hodgkin's disease. *J Nucl Med,* 43**,** 1018-27.

KRAK, N. C., BOELLAARD, R., HOEKSTRA, O. S., TWISK, J. W., HOEKSTRA, C. J. & LAMMERTSMA, A. A. 2005. Effects of ROI definition and reconstruction method on quantitative outcome and applicability in a response monitoring trial. *Eur J Nucl Med Mol Imaging,* 32**,** 294-301.

KRAK, N. C., VAN DER HOEVEN, J. J., HOEKSTRA, O. S., TWISK, J. W., VAN DER WALL, E. & LAMMERTSMA, A. A. 2003. Measuring [(18)F]FDG uptake in breast cancer

during chemotherapy: comparison of analytical methods. *Eur J Nucl Med Mol Imaging,* 30**,** 674-81.

KRAUSE, B. J., HERRMANN, K., WIEDER, H. & ZUM BUSCHENFELDE, C. M. 2009. F-18-FDG PET and F-18-FDG PET/CT for Assessing Response to Therapy in Esophageal Cancer. *Journal of Nuclear Medicine,* 50**,** 89s-96s.

KROEP, J. R., VAN GROENINGEN, C. J., CUESTA, M. A., CRAANEN, M. E., HOEKSTRA, O. S., COMANS, E. F., BLOEMENA, E., HOEKSTRA, C. J., GOLDING, R. P., TWISK, J. W., PETERS, G. J., PINEDO, H. M. & LAMMERTSMA, A. A. 2003. Positron emission tomography using 2-deoxy-2-[18F]-fluoro-D-glucose for response monitoring in locally advanced gastroesophageal cancer; a comparison of different analytical methods. *Mol Imaging Biol,* 5**,** 337-46.

KUHL, D. E., PHELPS, M. E., MARKHAM, C. H., METTER, E. J., RIEGE, W. H. & WINTER, J. 1982. Cerebral metabolism and atrophy in Huntington's disease determined by 18FDG and computed tomographic scan. *Ann Neurol,* 12**,** 425-34.

KURANI, A. S., XU, D. H., FURST, J. & RAICU, D. S. 2004. Co-occurrence matrices for volumetric data. *Proceedings of the Seventh IASTED International Conference on Computer Graphics and Imaging***,** 447-452.

LAMBIN, P., RIOS-VELAZQUEZ, E., LEIJENAAR, R., CARVALHO, S., VAN STIPHOUT, R. G., GRANTON, P., ZEGERS, C. M., GILLIES, R., BOELLARD, R., DEKKER, A. & AERTS, H. J. 2012. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer,* 48**,** 441-6.

LAMMERTSMA, A. A. 2001. Measurement of tumor response using [18F]-2-fluoro-2-deoxy-D-glucose and positron-emission tomography. *J Clin Pharmacol,* Suppl**,** 104S-106S.

LAMMERTSMA, A. A., HOEKSTRA, C. J., GIACCONE, G. & HOEKSTRA, O. S. 2006. How should we analyse FDG PET studies for monitoring tumour response? *Eur J Nucl Med Mol Imaging,* 33 Suppl 1**,** 16-21.

LANDIS, J. R. & KOCH, G. G. 1977. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics,* 33**,** 363-74.

LANGEN, K. J., BRAUN, U., ROTA KOPS, E., HERZOG, H., KUWERT, T., NEBELING, B. & FEINENDEGEN, L. E. 1993. The influence of plasma glucose levels on fluorine-18-fluorodeoxyglucose uptake in bronchial carcinomas. *J Nucl Med,* 34**,** 355-9.

LARSON, S. M., ERDI, Y., AKHURST, T., MAZUMDAR, M., MACAPINLAC, H. A., FINN, R. D., CASILLA, C., FAZZARI, M., SRIVASTAVA, N., YEUNG, H. W., HUMM, J. L., GUILLEM, J., DOWNEY, R., KARPEH, M., COHEN, A. E. & GINSBERG, R. 1999. Tumor Treatment Response Based on Visual and Quantitative Changes in Global Tumor Glycolysis Using PET-FDG Imaging. The Visual Response Score and the Change in Total Lesion Glycolysis. *Clin Positron Imaging,* 2**,** 159-171.

LARSON, S. M. & SCHWARTZ, L. H. 2006. F-18-FDG PET as a candidate for "Qualified Biomarker": Functional assessment of treatment response in oncology. *Journal of Nuclear Medicine,* 47**,** 901-903.

LECOMTE, R. 2009. Novel detector technology for clinical PET. *Eur J Nucl Med Mol Imaging,* 36 Suppl 1**,** S69-85.

LEE, H. Y., HYUN, S. H., LEE, K. S., KIM, B. T., KIM, J., SHIM, Y. M., AHN, M. J., KIM, T. S., YI, C. A. & CHUNG, M. J. 2010. Volume-Based Parameter of F-18-FDG PET/CT in Malignant Pleural Mesothelioma: Prediction of Therapeutic Response and Prognostic Implications. *Ann Surg Oncol,* 17**,** 2787-2794.

LEE, J. A. 2010. Segmentation of positron emission tomography images: Some recommendations for target delineation in radiation oncology. *Radiotherapy and Oncology,* 96**,** 302-307.

LEE, D. S., KANG, H., KIM, H., PARK, H., OH, J. S., LEE, J. S. & LEE, M. C. 2008. Metabolic connectivity by interregional correlation analysis using statistical parametric mapping (SPM) and FDG brain PET; methodological development and patterns of metabolic connectivity in adults. *Eur J Nucl Med Mol Imaging,* 35**,** 1681-91.

LEE, S. T., GHANEM, M., HERBERTSON, R. A., BERLANGIERI, S. U., BYRNE, A. J., TABONE, K., MITCHELL, P., KNIGHT, S. R., FEIGEN, M. & SCOTT, A. M. 2009. Prognostic Value of F-18-FDG PET/CT in Patients with Malignant Pleural Mesothelioma. *Molecular Imaging and Biology,* 11**,** 473-479.

LEE, T. T., EVERETT, D. L., SHU, H. K. G., JAHAN, T. M., ROACH, M., SPEIGHT, J. L., CAMERON, R. B., PHILLIPS, T. L., CHAN, A. & JABLONS, D. M. 2002. Radical pleurectomy/decortication and intraoperative radiotherapy followed by conformal radiation with or without chemotherapy for malignant pleural mesothelioma. *Journal of Thoracic and Cardiovascular Surgery,* 124**,** 1183-1189.

LERSKI, R. A., STRAUGHAN, K., SCHAD, L. R., BOYCE, D., BLUML, S. & ZUNA, I. 1993. Tissue Characterization by Magnetic-Resonance Spectroscopy and Imaging - Results of a Concerted Research-Project of the European-Economic-Community .8. Mr Image Texture Analysis - an Approach to Tissue Characterization. *Magn Reson Imaging,* 11**,** 873-887.

LEWELLEN, T. K. 2008. Recent developments in PET detector technology. *Phys Med Biol,* 53**,** R287-317.

LEWIS, J. S., MCCARTHY, D. W., MCCARTHY, T. J., FUJIBAYASHI, Y. & WELCH, M. J. 1999. Evaluation of 64Cu-ATSM in vitro and in vivo in a hypoxic tumor model. *J Nucl Med,* 40**,** 177-83.

LI, H., THORSTAD, W. L., BIEHL, K. J., LAFOREST, R., SU, Y., SHOGHI, K. I., DONNELLY, E. D., LOW, D. A. & LU, W. 2008. A novel PET tumor delineation method based on adaptive region-growing and dual-front active contours. *Med Phys,* 35**,** 3711-3721.

LIAO, S., PENNEY, B. C., ZHANG, H., SUZUKI, K. & PU, Y. 2012. Prognostic value of the quantitative metabolic volumetric measurement on 18F-FDG PET/CT in Stage IV nonsurgical small-cell lung cancer. *Acad Radiol,* 19**,** 69-77.

LIM, R., EATON, A., LEE, N. Y., SETTON, J., OHRI, N., RAO, S., WONG, R., FURY, M. & SCHODER, H. 2012. 18F-FDG PET/CT Metabolic Tumor Volume and Total Lesion Glycolysis Predict Outcome in Oropharyngeal Squamous Cell Carcinoma. *J Nucl Med,* 53**,** 1506-13.

LIN, C., ITTI, E., HAIOUN, C., PETEGNIEF, Y., LUCIANI, A., DUPUIS, J., PAONE, G., TALBOT, J. N., RAHMOUNI, A. & MEIGNAN, M. 2007. Early F-18-FDG PET for prediction of prognosis in patients with diffuse large B-cell lymphoma: SUV-based assessment versus visual analysis. *Journal of Nuclear Medicine,* 48**,** 1626-1632.

LINDHOLM, P., MINN, H., LESKINEN-KALLIO, S., BERGMAN, J., RUOTSALAINEN, U. & JOENSUU, H. 1993. Influence of the blood glucose concentration on FDG uptake in cancer--a PET study. *J Nucl Med,* 34**,** 1-6.

LITTLE, J. A., HILL, D. L. G. & HAWKES, D. J. 1997. Deformations Incorporating Rigid Structures. *Computer Vision and Image Understanding,* 66**,** 223–232.

LOCKHART, C. M., MACDONALD, L. R., ALESSIO, A. M., MCDOUGALD, W. A., DOOT, R. K. & KINAHAN, P. E. 2011. Quantifying and reducing the effect of calibration error on variability of PET/CT standardized uptake value measurements. *J Nucl Med,* 52**,** 218-24.

LODGE, M. A., CHAUDHRY, M. A. & WAHL, R. L. 2012. Noise Considerations for PET Quantification Using Maximum and Peak Standardized Uptake Value. *J Nucl Med,* 53**,** 1041-7.

LODGE, M. A., LUCAS, J. D., MARSDEN, P. K., CRONIN, B. F., O'DOHERTY, M. J. & SMITH, M. A. 1999. A PET study of 18FDG uptake in soft tissue masses. *Eur J Nucl Med,* 26**,** 22-30.

LONDON, K., CROSS, S., ONIKUL, E., DALLA-POZZA, L. & HOWMAN-GILES, R. 2011. 18F-FDG PET/CT in paediatric lymphoma: comparison with conventional imaging. *Eur J Nucl Med Mol Imaging,* 38**,** 274-84.

LONSDALE, M. N. & BEYER, T. 2010. Dual-modality PET/CT instrumentation-today and tomorrow. *Eur J Radiol,* 73**,** 452-60.

LOWE, V. J., DELONG, D. M., HOFFMAN, J. M. & COLEMAN, R. E. 1995. Optimum scanning protocol for FDG-PET evaluation of pulmonary malignancy. *J Nucl Med,* 36**,** 883-7.

LU, P. 2005. Staging and classification of lymphoma. *Semin Nucl Med,* 35**,** 160-164.

LUCIGNANI, G. 2008. PET imaging with hypoxia tracers: a must in radiation therapy. *Eur J Nucl Med Mol Imaging,* 35**,** 838-42.

LUCIGNANI, G. & DE PALMA, D. 2011. PET/CT in paediatric oncology: clinical usefulness and dosimetric concerns. *Eur J Nucl Med Mol Imaging,* 38**,** 179-184.

LUCIGNANI, G., PAGANELLI, G. & BOMBARDIERI, E. 2004. The use of standardized uptake values for assessing FDG uptake with PET in oncology: a clinical perspective. *Nucl Med Commun,* 25**,** 651-6.

MACMANUS, M., D'COSTA, I., EVERITT, S., ANDREWS, J., ACKERLY, T., BINNS, D., LAU, E., BALL, D., WEIH, L. & HICKS, R. J. 2007. Comparison of CT and positron emission tomography/CT coregistered images,in planning radical radiotherapy in patients with non-small-cell lung cancer. *Australas Radiol,* 51**,** 386-393.

MAHMOUD-GHONEIM, D., TOUSSAINT, G., CONSTANS, J. M. & DE CERTAINES, J. D. 2003. Three dimensional texture analysis in MRI: a preliminary evaluation in gliomas. *Magn Reson Imaging,* 21**,** 983-987.

MAISEY, M. N. 2002. Overview of clinical PET. *Br J Radiol,* 75 Spec No**,** S1-5.

MAK, V., DAVIES, E., PUTCHA, V., CHOODARI-OSKOOEI, B. & MOLLER, H. 2008. The epidemiology and treatment of mesothelioma in South East England 1985-2002. *Thorax,* 63**,** 160-166.

MARSDEN, P. & SUTCLIFFE-GOULDEN, J. 2000. Principles and technology of PET scanning. *Nucl Med Commun,* 21**,** 221-4.

MARSDEN, P. K. 2004. Quantification in PET: what is it? Can we do it? Do we need it? *Nucl Med Commun,* 25**,** 635-6.

MARTELLI, M., FERRERI, A. J., AGOSTINELLI, C., DI ROCCO, A., PFREUNDSCHUH, M. & PILERI, S. A. 2013. Diffuse large B-cell lymphoma. *Crit Rev Oncol Hematol*.

MCKELVEY, E. M., GOTTLIEB, J. A., WILSON, H. E., HAUT, A., TALLEY, R. W., STEPHENS, R., LANE, M., GAMBLE, J. F., JONES, S. E., GROZEA, P. N., GUTTERMAN, J., COLTMAN, C. & MOON, T. E. 1976. Hydroxyldaunomycin (Adriamycin) combination chemotherapy in malignant lymphoma. *Cancer,* 38**,** 1484-93.

MCQUAID, S. J., LAMBROU, T. & HUTTON, B. F. 2011. A novel method for incorporating respiratory-matched attenuation correction in the motion correction of cardiac PET-CT studies. *Phys Med Biol,* 56**,** 2903-15.

MEIGNAN, M., GALLAMINI, A. & HAIOUN, C. 2009. Report on the First International Workshop on Interim-PET-Scan in Lymphoma. *Leuk Lymphoma,* 50**,** 1257-60.

MEIGNAN, M., GALLAMINI, A. & HAIOUN, C. 2009. Report on the First International Workshop on Interim-PET-Scan in Lymphoma. *Leuk Lymphoma,* 50**,** 1257-60.

MEIGNAN, M., GALLAMINI, A., HAIOUN, C. & POLLIACK, A. 2010. Report on the Second International Workshop on interim positron emission tomography in lymphoma held in Menton, France, 8-9 April 2010. *Leuk Lymphoma,* 51**,** 2171-80.

MEIGNAN, M., GALLAMINI, A., ITTI, E., BARRINGTON, S., HAIOUN, C. & POLLIACK, A. 2012. Report on the Third International Workshop on Interim Positron Emission Tomography in Lymphoma held in Menton, France, 26-27 September 2011 and Menton 2011 consensus. *Leuk Lymphoma,* 53**,** 1876-81.

MEIGNAN, M., ITTI, E., GALLAMINI, A. & HAIOUN, C. 2009. Interim 18F-fluorodeoxyglucose positron emission tomography in diffuse large B-cell lymphoma: qualitative or quantitative interpretation--where do we stand? *Leuk Lymphoma,* 50**,** 1753-6.

MEIKLE, S. R., HUTTON, B. F., BAILEY, D. L., HOOPER, P. K. & FULHAM, M. J. 1994. Accelerated EM reconstruction in total-body PET: potential for improving tumour detectability. *Phys Med Biol,* 39**,** 1689-704.

MENDA, Y., BUSHNELL, D. L., MADSEN, M. T., MCLAUGHLIN, K., KAHN, D. & KERNSTINE, K. H. 2001. Evaluation of various corrections to the standardized uptake value for diagnosis of pulmonary malignancy. *Nucl Med Commun,* 22**,** 1077-81.

MERCER, J. R. 2007. Molecular imaging agents for clinical positron emission tomography in oncology other than fluorodeoxyglucose (FDG): applications, limitations and potential. *J Pharm Pharm Sci,* 10**,** 180-202.

MEYER, C. R., BOES, J. L., KIM, B., BLAND, P. H., ZASADNY, K. R., KISON, P. V., KORAL, K., FREY, K. A. & WAHL, R. L. 1997. Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin-plate spline warped geometric deformations. *Med Image Anal,* 1**,** 195-206.

MIKHAEEL, N. G. 2006. Use of FDG-PET to monitor response to chemotherapy and radiotherapy in patients with lymphomas. *Eur J Nucl Med Mol Imaging,* 33**,** S22-S26.

MIKHAEEL, N. G. 2009. Interim fluorodeoxyglucose positron emission tomography for early response assessment in diffuse large B cell lymphoma: where are we now? *Leuk Lymphoma,* 50**,** 1931-6.

MIKHAEEL, N. G., HUTCHINGS, M., FIELDS, P. A., O'DOHERTY, M. J. & TIMOTHY, A. R. 2005. FDG-PET after two to three cycles of chemotherapy predicts progression-free and overall survival in high-grade non-Hodgkin lymphoma. *Ann Oncol,* 16**,** 1514-23.

MILLER, A. B., HOOGSTRATEN, B., STAQUET, M. & WINKLER, A. 1981. Reporting results of cancer treatment. *Cancer,* 47**,** 207-14.

MILLER, T. R., PINKUS, E., DEHDASHTI, F. & GRIGSBY, P. W. 2003. Improved prognostic value of F-18-FDG PET using a simple visual analysis of tumor characteristics in patients with cervical cancer. *Journal of Nuclear Medicine,* 44**,** 192-197.

MINN, H., ZASADNY, K. R., QUINT, L. E. & WAHL, R. L. 1995. Lung cancer: reproducibility of quantitative measurements for evaluating 2-[F-18]-fluoro-2-deoxy-D-glucose uptake at PET. *Radiology,* 196**,** 167-73.

MINN, H., LESKINEN-KALLIO, S., LINDHOLM, P., BERGMAN, J., RUOTSALAINEN, U., TERAS, M. & HAAPARANTA, M. 1993. [18F]fluorodeoxyglucose uptake in tumors: kinetic vs. steady-state methods with reference to plasma insulin. *J Comput Assist Tomogr,* 17**,** 115-23.

MIYAZAKI, Y., NAWA, Y., MIYAGAWA, M., KOHASHI, S., NAKASE, K., YASUKAWA, M. & HARA, M. 2013. Maximum standard uptake value of (18)F-fluorodeoxyglucose positron emission tomography is a prognostic factor for progression-free survival of newly diagnosed patients with diffuse large B cell lymphoma. *Ann Hematol,* 92**,** 239-44.

MOCCIA, A. A., SCHAFF, K., HOSKINS, P., KLASA, R., SAVAGE, K. J., SHENKIER, T., GASCOYNE, R. D., CONNORS, J. M. & SEHN, L. H. 2009. R-CHOP with Etoposide Substituted for Doxorubicin (R-CEOP): Excellent Outcome in Diffuse Large B Cell Lymphoma for Patients with a Contraindication to Anthracyclines. *Blood,* 114**,** 170-170.

MOSES, W. W. 2011. Fundamental Limits of Spatial Resolution in PET. *Nucl Instrum Methods Phys Res A,* 648 Supplement 1**,** S236-S240.

MOSKOWITZ, C. H., SCHODER, H., TERUYA-FELDSTEIN, J., SIMA, C., IASONOS, A., PORTLOCK, C. S., STRAUS, D., NOY, A., PALOMBA, M. L., O'CONNOR, O. A., HORWITZ, S., WEAVER, S. A., MEIKLE, J. L., FILIPPA, D. A., CARAVELLI, J. F., HAMLIN, P. A. & ZELENETZ, A. D. 2010. Risk-Adapted Dose-Dense Immunochemotherapy Determined by Interim FDG-PET in Advanced-Stage Diffuse Large B-Cell Lymphoma. *Journal of Clinical Oncology,* 28**,** 1896-1903.

MUDIGONDA, N. R., RANGAYYAN, R. M. & DESAUTELS, J. E. L. 2000. Gradient and texture analysis for the classification of mammographic masses. *IEEE Trans Med Imaging,* 19**,** 1032-1043.

NAHMIAS, C. & WAHL, L. M. 2008. Reproducibility of standardized uptake value measurements determined by 18F-FDG PET in malignant tumors. *J Nucl Med,* 49**,** 1804-8.

NAKAMOTO, Y., TATSUMI, M., COHADE, C., OSMAN, M., MARSHALL, L. T. & WAHL, R. L. 2003. Accuracy of image fusion of normal upper abdominal organs visualized with PET/CT. *Eur J Nucl Med Mol Imaging,* 30**,** 597-602.

NAKAMOTO, Y., ZASADNY, K. R., MINN, H. & WAHL, R. L. 2002. Reproducibility of common semi-quantitative parameters for evaluating lung cancer glucose metabolism with positron emission tomography using 2-deoxy-2-[18F]fluoro-D-glucose. *Mol Imaging Biol,* 4**,** 171-8.

NECIB, H., DUSART, M., VANDERLINDEN, B. & BUVAT, I. 2008. Detection and characterization of the tumor change between two FDG PET scans using parametric imaging. *2008 Ieee International Symposium on Biomedical Imaging: From Nano to Macro, Vols 1-4***,** 21-24.

NECIB, H., GARCIA, C., WAGNER, A., VANDERLINDEN, B., EMONTS, P., HENDLISZ, A., FLAMEN, P. & BUVAT, I. 2011. Detection and characterization of tumor changes in 18F-FDG PET patient monitoring using parametric imaging. *J Nucl Med,* 52**,** 354-61.

NEHMEH, S. A., EL-ZEFTAWY, H., GRECO, C., SCHWARTZ, J., ERDI, Y. E., KIROV, A., SCHMIDTLEIN, C. R., GYAU, A. B., LARSON, S. M. & HUMM, J. L. 2009. An iterative technique to segment PET lesions using a Monte Carlo based mathematical model. *Med Phys,* 36**,** 4803-4809.

NEHMEH, S. A., ERDI, Y. E., PAN, T., YORKE, E., MAGERAS, G. S., ROSENZWEIG, K. E., SCHODER, H., MOSTAFAVI, H., SQUIRE, O., PEVSNER, A., LARSON, S. M. & HUMM, J. L. 2004. Quantitation of respiratory motion during 4D-PET/CT acquisition. *Med Phys,* 31**,** 1333-8.

NESTLE, U., KREMP, S., SCHAEFER-SCHULER, A., SEBASTIAN-WELSCH, C., HELLWIG, D., RUBE, C. & KIRSCH, C. M. 2005. Comparison of different methods for delineation of 18F-FDG PET-positive tissue for target volume definition in radiotherapy of patients with non-Small cell lung cancer. *J Nucl Med,* 46**,** 1342-8.

NESTLE, U., SCHAEFER-SCHULER, A., KREMP, S., GROESCHEL, A., HELLWIG, D., RUBE, C. & KIRSCH, C. M. 2007. Target volume definition for 18F-FDG PET-positive lymph nodes in radiotherapy of patients with non-small cell lung cancer. *Eur J Nucl Med Mol Imaging,* 34**,** 453-62.

NESTLE, U., WEBER, W., HENTSCHEL, M. & GROSU, A. L. 2009. Biological imaging in radiation therapy: role of positron emission tomography. *Phys Med Biol,* 54**,** R1-25.

NOMORI, H., WATANABE, K., OHTSUKA, T., NARUKE, T., SUEMASU, K. & UNO, K. 2005. Visual and semi quantitative analyses for F-18 fluorodeoxyglucose PET scanning in pulmonary nodules 1 cm to 3 cm in size. *Annals of Thoracic Surgery,* 79**,** 984-989.

NOWAK, A. K., BYRNE, M. J., WILLIAMSON, R., RYAN, G., SEGAL, A., FIELDING, D., MITCHELL, P., MUSK, A. W. & ROBINSON, B. W. 2002. A multicentre phase II study of cisplatin and gemcitabine for malignant mesothelioma. *Br J Cancer,* 87**,** 491-6.

NOWAK, A. K., FRANCIS, R. J., PHILLIPS, M. J., MILLWARD, M. J., VAN DER SCHAAF, A. A., BOUCEK, J., MUSK, A. W., MCCOY, M. J., SEGAL, A., ROBINS, P. & BYRNE, M. J. 2010. A novel prognostic model for malignant mesothelioma incorporating quantitative FDG-PET imaging with clinical parameters. *Clin Cancer Res,* 16**,** 2409-17.

OH, M. Y., OH, S. B., SEOUNG, H. G., KIM, J. H., KIM, S. M., KIM, T. K., SONG, M. K., SHIN, H. J. & CHUNG, J. S. 2012. Clinical significance of standardized uptake value and maximum tumor diameter in patients with primary extranodal diffuse large B cell lymphoma. *Korean J Hematol,* 47**,** 207-12.

OHTA, Y., SHRIDHAR, V., BRIGHT, R. K., KALEMKERIAN, G. P., DU, W., CARBONE, M., WATANABE, Y. & PASS, H. I. 1999. VEGF and VEGF type C play an important role in angiogenesis and lymphangiogenesis in human malignant mesothelioma tumours. *Br J Cancer,* 81**,** 54-61.

OKAZUMI, S., ISONO, K., ENOMOTO, K., KIKUCHI, T., OZAKI, M., YAMAMOTO, H., HAYASHI, H., ASANO, T. & RYU, M. 1992. Evaluation of liver tumors using fluorine-18-

fluorodeoxyglucose PET: characterization of tumor and assessment of effect of treatment. *J Nucl Med,* 33**,** 333-9.

OKEN, M. M., CREECH, R. H., TORMEY, D. C., HORTON, J., DAVIS, T. E., MCFADDEN, E. T. & CARBONE, P. P. 1982. Toxicity and response criteria of the Eastern Cooperative Oncology Group. *Am J Clin Oncol,* 5**,** 649-55.

O'SULLIVAN, F., ROY, S. & EARY, J. 2003. A statistical measure of tissue heterogeneity with application to 3D PET sarcoma data. *Biostatistics,* 4**,** 433-448.

O'SULLIVAN, F., ROY, S., O'SULLIVAN, J., VERNON, C. & EARY, J. 2005. Incorporation of tumor shape into an assessment of spatial heterogeneity for human sarcomas imaged with FDG-PET. *Biostatistics,* 6**,** 293-301.

OUKSILI, Z., TAUBER, C., NALIS, J., BATATIA, H., CASELLES, O. & COURBON, F. 2007. Accurate PET-PET registration to assess lung tumor evolution. *Biomedical Imaging: From Nano to Macro, 2007. ISBI 2007. 4th IEEE International Symposium***,** 732-735.

OVERGAARD, J. 2007. Hypoxic radiosensitization: adored and ignored. *J Clin Oncol,* 25**,** 4066-74.

PAN, T., MAWLAWI, O., NEHMEH, S. A., ERDI, Y. E., LUO, D., LIU, H. H., CASTILLO, R., MOHAN, R., LIAO, Z. & MACAPINLAC, H. A. 2005. Attenuation correction of PET images with respiration-averaged CT images in PET/CT. *J Nucl Med,* 46**,** 1481-7.

PANTALEO, M. A., NANNINI, M., MALEDDU, A., FANTI, S., AMBROSINI, V., NANNI, C., BOSCHI, S. & BIASCO, G. 2008. Conventional and novel PET tracers for imaging in oncology in the era of molecular therapy. *Cancer Treat Rev,* 34**,** 103-21.

PAPA, S., POPAT, S., SHAH, R., MCLENNAN, B., LAL, R., LANG-LAZDUNSKI, L., MARSDEN, P., VINEY, Z., LANDAU, D. & SPICER, J. 2012. A Phase 2 Study of Sorafenib after First Line Platinum Containing Combination Chemotherapy in Malignant Mesothelioma. *Journal of Thoracic Oncology,* 7**,** S84.

PAPATHANASSIOU, D., BRUNA-MURAILLE, C., LIEHN, J. C., NGUYEN, T. D. & CURE, H. 2009. Positron Emission Tomography in oncology: Present and future of PET and PET/CT. *Critical Reviews in Oncology Hematology,* 72**,** 239-254.

PAQUET, N., ALBERT, A., FOIDART, J. & HUSTINX, R. 2004. Within-patient variability of (18)F-FDG: standardized uptake values in normal tissues. *J Nucl Med,* 45**,** 784-8.

PARK, S. J., ROGERS, W. L. & CLINTHORNE, N. H. 2007. Design of a very high-resolution small animal PET scanner using a silicon scatter detector insert. *Phys Med Biol,* 52**,** 4653-77.

PARK, J. O., LEE, S. I., SONG, S. Y., KIM, K., KIM, W. S., JUNG, C. W., PARK, Y. S., IM, Y. H., KANG, W. K., LEE, M. H., LEE, K. S. & PARK, K. 2003. Measuring response in solid tumors: comparison of RECIST and WHO response criteria. *Japanese Journal of Clinical Oncology,* 33**,** 533-7.

PATEL, C. N., GOLDSTONE, A. R., CHOWDHURY, F. U. & SCARSBROOK, A. F. 2010. FDG PET/CT in oncology: "raising the bar". *Clin Radiol,* 65**,** 522-35.

PAULINO, A. C. & JOHNSTONE, P. A. S. 2004. FDG-PET in radiotherapy treatment planning: Pandora's box? *International Journal of Radiation Oncology Biology Physics,* 59**,** 4-5.

PELUCCHI, C., MALVEZZI, M., LA VECCHIA, C., LEVI, F., DECARLI, A. & NEGRI, E. 2004. The Mesothelioma epidemic in Western Europe: an update. *Br J Cancer,* 90**,** 1022-4.

PENG, B. H. & LEVIN, C. S. 2010. Recent development in PET instrumentation. *Curr Pharm Biotechnol,* 11**,** 555-71.

PETO, J., DECARLI, A., LA VECCHIA, C., LEVIS, F. & NEGRI, E. 1999. The European mesothelioma epidemic. *Br J Cancer,* 79**,** 666-672.

PFREUNDSCHUH, M., KUHNT, E., TRUMPER, L., OSTERBORG, A., TRNENY, M., SHEPHERD, L., GILL, D. S., WALEWSKI, J., PETTENGELL, R., JAEGER, U., ZINZANI, P. L., SHPILBERG, O., KVALOY, S., DE NULLY BROWN, P., STAHEL, R., MILPIED, N., LOPEZ-GUILLERMO, A., POESCHEL, V., GRASS, S., LOEFFLER, M. & MURAWSKI, N. 2011. CHOP-like chemotherapy with or without rituximab in young patients with good-prognosis

diffuse large-B-cell lymphoma: 6-year results of an open-label randomised study of the MabThera International Trial (MInT) Group. *Lancet Oncol,* 12**,** 1013-22.

PHELPS, M. E., HOFFMAN, E. J., MULLANI, N. A. & TER-POGOSSIAN, M. M. 1975. Application of annihilation coincidence detection to transaxial reconstruction tomography. *J Nucl Med,* 16**,** 210-24.

PHELPS, M. E., HUANG, S. C., HOFFMAN, E. J., SELIN, C., SOKOLOFF, L. & KUHL, D. E. 1979. Tomographic measurement of local cerebral glucose metabolic rate in humans with (F-18)2-fluoro-2-deoxy-D-glucose: validation of method. *Ann Neurol,* 6**,** 371-88.

PIEN, H. H., FISCHMAN, A. J., THRALL, J. H. & SORENSEN, A. G. 2005. Using imaging biomarkers to accelerate drug development and clinical trials. *Drug Discovery Today,* 10**,** 259-66.

PORTWINE, C., MARRIOTT, C. & BARR, R. D. 2010. PET Imaging for Pediatric Oncology: An Assessment of the Evidence. *Pediatr Blood Cancer,* 55**,** 1048-1061.

PREGNO, P., CHIAPPELLA, A., BELLO, M., BOTTO, B., FERRERO, S., FRANCESCHETTI, S., GIUNTA, F., LADETTO, M., LIMERUTTI, G., MENGA, M., NICOLOSI, M., PRIOLO, G., PUCCINI, B., RIGACCI, L., SALVI, F., VAGGELLI, L., PASSERA, R., BISI, G. & VITOLO, U. 2012. Interim 18-FDG-PET/CT failed to predict the outcome in diffuse large B-cell lymphoma patients treated at the diagnosis with rituximab-CHOP. *Blood,* 119**,** 2066-73.

QUARANTELLI, M., BERKOUK, K., PRINSTER, A., LANDEAU, B., SVARER, C., BALKAY, L., ALFANO, B., BRUNETTI, A., BARON, J. C. & SALVATORE, M. 2004. Integrated software for the analysis of brain PET/SPECT studies with partial-volume-effect correction. *J Nucl Med,* 45**,** 192-201.

RAICHLE, M. E., PERLMUTTER, J. S. & FOX, P. T. 1984. Parkinson's disease: metabolic and pharmacological approaches with positron emission tomography. *Ann Neurol,* 15 Suppl**,** S131-2.

RANGAYYAN, R. M., EL-FARAMAWY, N. M., DESAUTELS, J. E. L. & ALIM, O. A. 1997. Measures of acutance and shape for classification of breast tumors. *IEEE Trans Med Imaging,* 16**,** 799-810.

RASEY, J. S., KOH, W. J., EVANS, M. L., PETERSON, L. M., LEWELLEN, T. K., GRAHAM, M. M. & KROHN, K. A. 1996. Quantifying regional hypoxia in human tumors with positron emission tomography of [18F]fluoromisonidazole: a pretherapy study of 37 patients. *Int J Radiat Oncol Biol Phys,* 36**,** 417-28.

REIVICH, M., KUHL, D., WOLF, A., GREENBERG, J., PHELPS, M., IDO, T., CASELLA, V., FOWLER, J., HOFFMAN, E., ALAVI, A., SOM, P. & SOKOLOFF, L. 1979. The [18F]fluorodeoxyglucose method for the measurement of local cerebral glucose utilization in man. *Circ Res,* 44**,** 127-37.

RESKE, S. N. & DEISENHOFER, S. 2006. Is 3'-deoxy-3'-(18)F-fluorothymidine a better marker for tumour response than (18)F-fluorodeoxyglucose? *Eur J Nucl Med Mol Imaging,* 33 Suppl 1**,** 38-43.

ROBERTS, G. H. 1970. Diffuse pleural mesothelioma. A clinical and pathological study. *British journal of diseases of the chest,* 64**,** 201-11.

ROBINSON, B. W. S. & LAKE, R. A. 2005. Medical progress - Advances in malignant mesothelioma. *New England Journal of Medicine,* 353**,** 1591-1603.

ROBINSON, B. W. S., MUSK, A. W. & LAKE, R. A. 2005. Malignant mesothelioma. *Lancet,* 366**,** 397-408.

ROEDL, J. B., COLEN, R. R., HOLALKERE, N. S., FISCHMAN, A. J., CHOI, N. C. & BLAKE, M. A. 2008. Adenocarcinomas of the esophagus: response to chemoradiotherapy is associated with decrease of metabolic tumor volume as measured on PET-CT. Comparison to histopathologic and clinical response evaluation. *Radiother Oncol,* 89**,** 278-86.

ROHREN, E. M., TURKINGTON, T. G. & COLEMAN, R. E. 2004. Clinical applications of PET in oncology. *Radiology,* 231**,** 305-32.

ROUSSET, O. G., MA, Y. & EVANS, A. C. 1998. Correction for partial volume effects in PET: principle and validation. *J Nucl Med,* 39**,** 904-11.

RUECKERT, D., SONODA, L. I., HAYES, C., HILL, D. L., LEACH, M. O. & HAWKES, D. J. 1999. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans Med Imaging,* 18**,** 712-21.

RUSCH, V. W. 1995. Clinical-Features and Current Treatment of Diffuse Malignant Pleural Mesothelioma. *Lung Cancer,* 12**,** S127-S146.

S.C, S., M.E., C. & E.J., H. 1990. Measuring PET Scanner Sensitivity: Relating Countrates to Image Signal-to-Noise Ratios using Noise Equivalent Counts. *IEEE Transactions on Nuclear Science,* 37**,** 783-788.

SADATO, N., TSUCHIDA, T., NAKAUMRA, S., WAKI, A., UEMATSU, H., TAKAHASHI, N., HAYASHI, N., YONEKURA, Y. & ISHII, Y. 1998. Non-invasive estimation of the net influx constant using the standardized uptake value for quantification of FDG uptake of tumours. *Eur J Nucl Med,* 25**,** 559-64.

SAFAR, V., DUPUIS, J., ITTI, E., JARDIN, F., FRUCHART, C., BARDET, S., VERA, P., COPIE-BERGMAN, C., RAHMOUNI, A., TILLY, H., MEIGNAN, M. & HAIOUN, C. 2012. Interim [18F]fluorodeoxyglucose positron emission tomography scan in diffuse large B-cell lymphoma treated with anthracycline-based chemotherapy plus rituximab. *J Clin Oncol,* 30**,** 184-90.

SCHAEFER, N. G., VEIT-HAIBACH, P., SOYKA, J. D., STEINERT, H. C. & STAHEL, R. A. 2012. Continued pemetrexed and platin-based chemotherapy in patients with malignant pleural mesothelioma (MPM): value of 18F-FDG-PET/CT. *Eur J Radiol,* 81**,** e19-25.

SCHIEPERS, C. & DAHLBOM, M. 2011. Molecular imaging in oncology: the acceptance of PET/CT and the emergence of MR/PET imaging. *Eur Radiol,* 21**,** 548-554.

SCHLEYER, D. J. 2004. PET Tracers and Radiochemistry. *Ann Acad Med Singapore,* 33**,** 146-154.

SCHNABEL, J. A., RUECKERT, D., QUIST, M., BLACKALL, J. M., CASTELLANO-SMITH, A. D., HARTKENS, T., PENNEY, G. P., HALL, W. A., LIU, H., TRUWIT, C. L., GERRITSEN,

F. A., HILL, D. L. G. & HAWKES, D. J. 2001. A Generic Framework for Non-rigid Registration Based on Non-uniform Multi-level Free-Form Deformations. *MICCI***,** 573-581.

SCHODER, H., FURY, M., LEE, N. & KRAUS, D. 2009. PET Monitoring of Therapy Response in Head and Neck Squamous Cell Carcinoma. *Journal of Nuclear Medicine,* 50**,** 74s-88s.

SCHWAIGER, M., ZIEGLER, S. & NEKOLLA, S. G. 2005. PET/CT: Challenge for nuclear cardiology. *Journal of Nuclear Medicine,* 46**,** 1664-1678.

SCHWARZ, J. K., GRIGSBY, P. W., DEHDASHTI, F. & DELBEKE, D. 2009. The Role of F-18-FDG PET in Assessing Therapy Response in Cancer of the Cervix and Ovaries. *Journal of Nuclear Medicine,* 50**,** 64s-73s.

SEHN, L. H., BERRY, B., CHHANABHAI, M., FITZGERALD, C., GILL, K., HOSKINS, P., KLASA, R., SAVAGE, K. J., SHENKIER, T., SUTHERLAND, J., GASCOYNE, R. D. & CONNORS, J. M. 2007. The revised International Prognostic Index (R-IPI) is a better predictor of outcome than the standard IPI for patients with diffuse large B-cell lymphoma treated with R-CHOP. *Blood,* 109**,** 1857-61.

SHANKAR, L. K., HOFFMAN, J. M., BACHARACH, S., GRAHAM, M. M., KARP, J., LAMMERTSMA, A. A., LARSON, S., MANKOFF, D. A., SIEGEL, B. A., DEN ABBEELE, A., YAP, J. & SULLIVAN, D. 2006. Consensus recommendations for the use of F-18-FDG PET as an indicator of therapeutic response in patients in national cancer institute trials. *Journal of Nuclear Medicine,* 47**,** 1059-1066.

SHARIF, M. S., ABBOD, M., AMIRA, A. & ZAIDI, H. 2010. Artificial Neural Network-Based System for PET Volume Segmentation. *Int J Biomed Imaging,* 2010.

SHARIF, S., ZAHID, I., ROUTLEDGE, T. & SCARCI, M. 2011. Does positron emission tomography offer prognostic information in malignant pleural mesothelioma? *Interact Cardiovasc Thorac Surg,* 12**,** 806-11.

SHARMA, P., GUPTA, A., PATEL, C., BAKHSHI, S., MALHOTRA, A. & KUMAR, R. 2012. Pediatric lymphoma: metabolic tumor burden as a quantitative index for treatment response evaluation. *Ann Nucl Med,* 26**,** 58-66.

SHEKHAR, R., WALIMBE, V., RAJA, S., ZAGRODSKY, V., KANVINDE, M., WU, G. & BYBEL, B. 2005. Automated 3-dimensional elastic registration of whole-body PET and CT from separate or combined scanners. *J Nucl Med,* 46**,** 1488-96.

SHEPP, L. A. & VARDI, Y. 1982. Maximum likelihood reconstruction for emission tomography. *IEEE Trans Med Imaging,* 1**,** 113-22.

SHIELDS, A. F., GRIERSON, J. R., DOHMEN, B. M., MACHULLA, H. J., STAYANOFF, J. C., LAWHORN-CREWS, J. M., OBRADOVICH, J. E., MUZIK, O. & MANGNER, T. J. 1998. Imaging proliferation in vivo with [F-18]FLT and positron emission tomography. *Nat Med,* 4**,** 1334-6.

SHIELDS, A. F., GRIERSON, J. R., KOZAWA, S. M. & ZHENG, M. 1996. Development of labeled thymidine analogs for imaging tumor proliferation. *Nucl Med Biol,* 23**,** 17-22.

SOKOLOFF, L., REIVICH, M., KENNEDY, C., DES ROSIERS, M. H., PATLAK, C. S., PETTIGREW, K. D., SAKURADA, O. & SHINOHARA, M. 1977. The [14C]deoxyglucose method for the measurement of local cerebral glucose utilization: theory, procedure, and normal values in the conscious and anesthetized albino rat. *J Neurochem,* 28**,** 897-916.

SOMER, E. J., BENATAR, N. A., O'DOHERTY, M. J., SMITH, M. A. & MARSDEN, P. K. 2007. Use of the CT component of PET-CT to improve PET-MR registration: demonstration in soft-tissue sarcoma. *Phys Med Biol,* 52**,** 6991-7006.

SONG, M. K., CHUNG, J. S., SHIN, H. J., LEE, S. M., LEE, S. E., LEE, H. S., LEE, G. W., KIM, S. J. & CHUNG, D. S. 2012. Clinical significance of metabolic tumor volume by PET/CT in stages II and III of diffuse large B cell lymphoma without extranodal site involvement. *Ann Hematol,* 91**,** 697-703.

SONG, M. K., CHUNG, J. S., SHIN, H. J., MOON, J. H., LEE, J. O., LEE, H. S., LEE, S. M., LEE, G. W., LEE, S. E. & KIM, S. J. 2012. Prognostic value of metabolic tumor volume on PET / CT in primary gastrointestinal diffuse large B cell lymphoma. *Cancer Science,* 103**,** 477-82.

SORET, M., BACHARACH, S. L. & BUVAT, I. 2007. Partial-volume effect in PET tumor imaging. *J Nucl Med,* 48**,** 932-45.

SORGER, D., PATT, M., KUMAR, P., WIEBE, L. I., BARTHEL, H., SEESE, A., DANNENBERG, C., TANNAPFEL, A., KLUGE, R. & SABRI, O. 2003. [18F]Fluoroazomycinarabinofuranoside (18FAZA) and [18F]Fluoromisonidazole (18FMISO): a comparative study of their selective uptake in hypoxic cells and PET imaging in experimental rat tumors. *Nucl Med Biol,* 30**,** 317-26.

SPAEPEN, K., STROOBANTS, S., DUPONT, P., VANDENBERGHE, P., THOMAS, J., DE GROOT, T., BALZARINI, J., DE WOLF-PEETERS, C., MORTELMANS, L. & VERHOEF, G. 2002. Early restaging positron emission tomography with (18)F-fluorodeoxyglucose predicts outcome in patients with aggressive non-Hodgkin's lymphoma. *Ann Oncol,* 13**,** 1356-63.

SPANOUDAKI, V. & LEVIN, C. S. 2010. Photo-detectors for time of flight positron emission tomography (ToF-PET). *Sensors (Basel),* 10**,** 10484-505.

STAHL, A., OTT, K., SCHWAIGER, M. & WEBER, W. A. 2004. Comparison of different SUV-based methods for monitoring cytotoxic therapy with FDG PET. *Eur J Nucl Med Mol Imaging,* 31**,** 1471-8.

STEINERT, H. C., DELLEA, M. M. S., BURGER, C. & STAHEL, R. 2005. Therapy response evaluation in malignant pleural mesothelioma with integrated PET-CT imaging. *Lung Cancer,* 49**,** S33-S35.

STRASTERS, K. C., LITTLE, J. A., BUURMAN, J., HILL, D. L. G. & HAWKES, D. J. 1997. Anatomical landmark image registration: Validation and comparison. *Lecture Notes in Computer Science,* 1205**,** 161-170.

STUDHOLME, C., HILL, D. L. G. & HAWKES, D. J. 1999. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition,* 32**,** 71-86.

SUGARBAKER, D. J., FLORES, R. M., JAKLITSCH, M. T., RICHARDS, W. G., STRAUSS, G. M., CARSON, J. M., DECAMP, M. M., SWANSON, S. J., BUENO, R., LUKANICH, J. M., BALDINI, E. H. & MENTZER, S. J. 1999. Resection margins, extrapleural nodal status, and cell type determine postoperative long-term survival in trimodality therapy of malignant pleural mesothelioma: Results in 183 patients. *Journal of Thoracic and Cardiovascular Surgery,* 117**,** 54-63.

SUGARBAKER, D. J., GARCIA, J. P., RICHARDS, W. G., HARPOLE, D. H., HEALYBALDINI, E., DECAMP, M. M., MENTZER, S. J., LIPTAY, M. J., STRAUSS, G. M. & SWANSON, S. J. 1996. Extrapleural pneumonectomy in the multimodality therapy of malignant pleural mesothelioma - Results in 120 consecutive patients. *Ann Surg,* 224**,** 288-294.

SUGARBAKER, D. J., JAKLITSCH, M. T., BUENO, R., RICHARDS, W., LUKANICH, J., MENTZER, S. J., COLSON, Y., LINDEN, P., CHANG, M., CAPALBO, L., OLDREAD, E., NERAGI-MIANDOAB, S., SWANSON, S. J. & ZELLOS, L. S. 2004. Prevention, early detection, and management of complications after 328 consecutive extrapleural pneumonectomies. *Journal of Thoracic and Cardiovascular Surgery,* 128**,** 138-146.

SURTI, S., SCHEUERMANN, J., EL FAKHRI, G., DAUBE-WITHERSPOON, M. E., LIM, R., ABI-HATEM, N., MOUSSALLEM, E., BENARD, F., MANKOFF, D. & KARP, J. S. 2011. Impact of Time-of-Flight PET on Whole-Body Oncologic Studies: A Human Observer Lesion Detection and Localization Study. *Journal of Nuclear Medicine,* 52**,** 712-719.

SWEETENHAM, J. W. 2005. Diffuse large B-cell lymphoma: risk stratification and management of relapsed disease. *Hematology Am Soc Hematol Educ Program***,** 252-9.

TAI, Y. F. & PICCINI, P. 2004. Applications of positron emission tomography (PET) in neurology. *J Neurol Neurosurg Psychiatry,* 75**,** 669-76.

TAKAHASHI, Y., ORIUCHI, N., OTAKE, H., ENDO, K. & MURASE, K. 2008. Variability of lesion detectability and standardized uptake value according to the acquisition procedure and reconstruction among five PET scanners. *Ann Nucl Med,* 22**,** 543-548.

TAKIKAWA, S., DHAWAN, V., SPETSIERIS, P., ROBESON, W., CHALY, T., DAHL, R., MARGOULEFF, D. & EIDELBERG, D. 1993. Noninvasive quantitative fluorodeoxyglucose PET studies with an estimated input function derived from a population-based arterial blood curve. *Radiology,* 188**,** 131-6.

TARANTOLA, G., ZITO, F. & GERUNDINI, P. 2003. PET instrumentation and reconstruction algorithms in whole-body applications. *Journal of Nuclear Medicine,* 44**,** 756-769.

TEO, B. K., SEO, Y., BACHARACH, S. L., CARRASQUILLO, J. A., LIBUTTI, S. K., SHUKLA, H., HASEGAWA, B. H., HAWKINS, R. A. & FRANC, B. L. 2007. Partial-volume correction in PET: validation of an iterative postreconstruction method with phantom and patient data. *J Nucl Med,* 48**,** 802-10.

TERASAWA, T., LAU, J., BARDET, S., COUTURIER, O., HOTTA, T., HUTCHINGS, M., NIHASHI, T. & NAGAI, H. 2009. Fluorine-18-fluorodeoxyglucose positron emission tomography for interim response assessment of advanced-stage Hodgkin's lymphoma and diffuse large B-cell lymphoma: a systematic review. *J Clin Oncol,* 27**,** 1906-14.

TER-POGOSSIAN, M. M., PHELPS, M. E., HOFFMAN, E. J. & MULLANI, N. A. 1975. A positron-emission transaxial tomograph for nuclear imaging (PETT). *Radiology,* 114**,** 89-98.

TESAR, L., SHIMIZU, A., SMUTEK, D., KOBATAKE, H. & NAWANO, S. 2008. Medical image analysis of 3D CT images based on extension of Haralick texture features. *Computerized Medical Imaging and Graphics,* 32**,** 513-520.

THERASSE, P., ARBUCK, S. G., EISENHAUER, E. A., WANDERS, J., KAPLAN, R. S., RUBINSTEIN, L., VERWEIJ, J., VAN GLABBEKE, M., VAN OOSTEROM, A. T., CHRISTIAN, M. C. & GWYTHER, S. G. 2000. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer,

National Cancer Institute of the United States, National Cancer Institute of Canada. *J Natl Cancer Inst,* 92**,** 205-16.

THIE, J. A., HUBNER, K. F., ISIDORO, F. P. & SMITH, G. T. 2007. A weight index for the standardized uptake value in 2-deoxy-2-[F-18]fluoro-D-glucose-positron emission tomography. *Mol Imaging Biol,* 9**,** 91-8.

THIESSE, P., OLLIVIER, L., DI STEFANO-LOUINEAU, D., NEGRIER, S., SAVARY, J., PIGNARD, K., LASSET, C. & ESCUDIER, B. 1997. Response rate accuracy in oncology trials: reasons for interobserver variability. Groupe Francais d'Immunotherapie of the Federation Nationale des Centres de Lutte Contre le Cancer. *J Clin Oncol,* 15**,** 3507-14.

TIXIER, F., HATT, M., LE REST, C. C., LE POGAM, A., CORCOS, L. & VISVIKIS, D. 2012. Reproducibility of tumor uptake heterogeneity characterization through textural feature analysis in 18F-FDG PET. *J Nucl Med,* 53**,** 693-700.

TIXIER, F., LE REST, C. C., HATT, M., ALBARGHACH, N., PRADIER, O., METGES, J. P., CORCOS, L. & VISVIKIS, D. 2011. Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer. *J Nucl Med,* 52**,** 369-78.

TOMASI, G. & ROSSO, L. 2012. PET imaging: implications for the future of therapy monitoring with PET/CT in oncology. *Curr Opin Pharmacol,* 12**,** 569-75.

TOMASI, G., TURKHEIMER, F. & ABOAGYE, E. 2012. Importance of Quantification for the Analysis of PET Data in Oncology: Review of Current Methods and Trends for the Future. *Molecular Imaging and Biology,* 14**,** 131-146.

TORIZUKA, T., NAKAMURA, F., KANNO, T., FUTATSUBASHI, M., YOSHIKAWA, E., OKADA, H., KOBAYASHI, M. & OUCHI, Y. 2004. Early therapy monitoring with FDG-PET in aggressive non-Hodgkin's lymphoma and Hodgkin's lymphoma. *Eur J Nucl Med Mol Imaging,* 31**,** 22-28.

TOWNSEND, D. W. 2004. Physical principles and technology of clinical PET imaging. *Ann Acad Med Singapore,* 33**,** 133-45.

TRILLET-LENOIR, V., FREYER, G., KAEMMERLEN, P., FOND, A., PELLET, O., LOMBARD-BOHAS, C., GAUDIN, J. L., LLEDO, G., MACKIEWICZ, R., GOUTTEBEL, M. C., MOINDROT, H., BOYER, J. D., CHASSIGNOL, L., STREMSDOERFER, N., DESSEIGNE, F., MOREAU, J. M., HEDELIUS, F., MORAILLON, A., CHAPUIS, F., BLEUSE, J. P., BARBIER, Y., HEILMANN, M. O. & VALETTE, P. J. 2002. Assessment of tumour response to chemotherapy for metastatic colorectal cancer: accuracy of the RECIST criteria. *Br J Radiol,* 75**,** 903-8.

TSAI, F., CHANG, C. K., RAU, J. Y., LIN, T. H. & LIU, G. R. 2007. 3D computation of gray level co-occurrence in hyperspectral image cubes. *Energy Minimization Methods in Computer Vision and Pattern Recognition, Proceedings,* 4679**,** 429-440.

TUMA, R. S. 2006. Sometimes size doesn't matter: reevaluating RECIST and tumor response rate endpoints. *J Natl Cancer Inst,* 98**,** 1272-4.

UEMURA, K., TOYAMA, H., IKOMA, Y., ODA, K., SENDA, M. & UCHIYAMA, A. 1999. Correction of partial volume effect on rate constant estimation in compartment model analysis of dynamic PET study. *1999 Ieee Nuclear Science Symposium - Conference Record, Vols 1-3***,** 1128-1131.

VAIDYA, M., CREACH, K. M., FRYE, J., DEHDASHTI, F., BRADLEY, J. D. & EL NAQA, I. 2012. Combined PET/CT image characteristics for radiotherapy tumor response in lung cancer. *Radiotherapy and Oncology,* 102**,** 239-245.

VAN DALEN, J. A., HOFFMANN, A. L., DICKEN, V., VOGEL, W. V., WIERING, B., RUERS, T. J., KARSSEMEIJER, N. & OYEN, W. J. G. 2007. A novel iterative method for lesion delineation and volumetric quantification with FDG PET. *Nucl Med Commun,* 28**,** 485–493.

VAN KLAVEREN, R. J., AERTS, J. G. J. V., DE BRUIN, H., GIACCONE, G., MANEGOLD, C. & VAN MEERBEECK, J. P. 2004. Inadequacy of the RECIST criteria for response evaluation in patients with malignant pleural mesothelioma. *Lung Cancer,* 43**,** 63-69.

VAN VELDEN, F. H., CHEEBSUMON, P., YAQUB, M., SMIT, E. F., HOEKSTRA, O. S., LAMMERTSMA, A. A. & BOELLAARD, R. 2011. Evaluation of a cumulative SUV-volume histogram method for parameterizing heterogeneous intratumoural FDG uptake in non-small cell lung cancer PET studies. *Eur J Nucl Med Mol Imaging,* 38**,** 1636-47.

VAN VELDEN, F. H. P., LOECKX, D., VELASQUEZ, L., HAYES, W., HOETJES, N. J., BOELLAARD, R. & NUYTS, J. 2010. Effect of Rigid and Non-Rigid Image Registration on Test-Retest Repeatability of Quantitative Measures Derived from FDG PET/CT Oncology Studies. *Eur J Nucl Med Mol Imaging,* 37**,** S262-S262.

VANDERHOEK, M., PERLMAN, S. B. & JERAJ, R. 2012. Impact of the definition of peak standardized uptake value on quantification of treatment response. *J Nucl Med,* 53**,** 4-11.

VAUPEL, P. & MAYER, A. 2007. Hypoxia in cancer: significance and impact on clinical outcome. *Cancer Metastasis Rev,* 26**,** 225-39.

VEIT, P., RUEHM, S., KUEHL, H., STERGAR, H., MUELLER, S., BOCKISCH, A. & ANTOCH, G. 2006. Lymph node staging with dual-modality PET/CT: Enhancing the diagnostic accuracy in oncology. *Eur J Radiol,* 58**,** 383-389.

VEIT-HAIBACH, P., SCHAEFER, N. G., STEINERT, H. C., SOYKA, J. D., SEIFERT, B. & STAHEL, R. A. 2010. Combined FDG-PET/CT in response evaluation of malignant pleural mesothelioma. *Lung Cancer,* 67**,** 311-317.

VISVIKIS, D., CHEZE-LEREST, C., COSTA, D. C., BOMANJI, J., GACINOVIC, S. & ELL, P. J. 2001. Influence of OSEM and segmented attenuation correction in the calculation of standardised uptake values for [18F]FDG PET. *Eur J Nucl Med,* 28**,** 1326-35.

VISVIKIS, D., TURZO, A., BIZAIS, Y. & CHEZE-LE REST, C. 2004. Technology related parameters affecting quantification in positron emission tomography imaging. *Nucl Med Commun,* 25**,** 637-41.

VOGELZANG, N. J., RUSTHOVEN, J. J., SYMANOWSKI, J., DENHAM, C., KAUKEL, E., RUFFIE, P., GATZEMEIER, U., BOYER, M., EMRI, S., MANEGOLD, C., NIYIKIZA, C. & PAOLETTI, P. 2003. Phase III study of pemetrexed in combination with cisplatin versus cisplatin alone in patients with malignant pleural mesothelioma. *Journal of Clinical Oncology,* 21**,** 2636-2644.

WAHL, R. L., JACENE, H., KASAMON, Y. & LODGE, M. A. 2009. From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors. *J Nucl Med,* 50 Suppl 1**,** 122S-50S.

WAHL, R. L., ZASADNY, K., HELVIE, M., HUTCHINS, G. D., WEBER, B. & CODY, R. 1993. Metabolic monitoring of breast cancer chemohormonotherapy using positron emission tomography: initial evaluation. *J Clin Oncol,* 11**,** 2101-11.

WANG, Z. Y., REDDY, G. P., GOTWAY, M. B., HIGGINS, C. B., JABLONS, D. M., RAMASWAMY, M., HAWKINS, R. A. & WEBB, W. R. 2004. Malignant pleural mesothelioma: Evaluation with CT, MR imaging, and PET. *Radiographics,* 24**,** 105-119.

WARBURG, O. 1956. On the origin of cancer cells. *Science,* 123**,** 309-14.

WEBER, W. A. 2005. PET for response assessment in oncology: radiotherapy and chemotherapy. *British Journal of Radiology,* 78**,** 42-49.

WEIGERT, M., PIETRZYK, U., MULLER, S., PALM, C. & BEYER, T. 2008. Whole-body PET/CT imaging: combining software- and hardware-based co-registration. *Zeitschrift Fur Medizinische Physik,* 18**,** 59-66.

WESTERTERP, M., PRUIM, J., OYEN, W., HOEKSTRA, O., PAANS, A., VISSER, E., VAN LANSCHOT, J., SLOOF, G. & BOELLAARD, R. 2007. Quantification of FDG PET studies

using standardised uptake values in multi-centre trials: effects of image reconstruction, resolution and ROI definition parameters. *Eur J Nucl Med Mol Imaging,* 34**,** 392-404.

WIEBE, L. I. 2004. PET radiopharmaceuticals for metabolic imaging in oncology. *Pet and Molecular Imaging: State of the Art and Future Perspectives,* 1264**,** 53-76.

WIEDER, H. A., BEER, A. J., LORDICK, F., OTT, K., FISCHER, M., RUMMENY, E. J., ZIEGLER, S., SIEWER, J. R., SCHWAIGER, M. & WEBER, W. A. 2005. Comparison of changes in tumor metabolic activity and tumor size during chemotherapy of adenocarcinomas of the esophagogastric junction. *J Nucl Med,* 46**,** 2029-34.

WILHELM, S. M., CARTER, C., TANG, L., WILKIE, D., MCNABOLA, A., RONG, H., CHEN, C., ZHANG, X., VINCENT, P., MCHUGH, M., CAO, Y., SHUJATH, J., GAWLAK, S., EVELEIGH, D., ROWLEY, B., LIU, L., ADNANE, L., LYNCH, M., AUCLAIR, D., TAYLOR, I., GEDRICH, R., VOZNESENSKY, A., RIEDL, B., POST, L. E., BOLLAG, G. & TRAIL, P. A. 2004. BAY 43-9006 exhibits broad spectrum oral antitumor activity and targets the RAF/MEK/ERK pathway and receptor tyrosine kinases involved in tumor progression and angiogenesis. *Cancer Research,* 64**,** 7099-109.

WONG, J. C., STUDHOLME, C., HAWKES, D. J. & MAISEY, M. N. 1997. Evaluation of the limits of visual detection of image misregistration in a brain fluorine-18 fluorodeoxyglucose PET-MRI study. *Eur J Nucl Med,* 24**,** 642-50.

WOODARD, H. Q., BIGLER, R. E. & FREED, B. 1975. Letter: Expression of tissue isotope distribution. *J Nucl Med,* 16**,** 958-9.

WU, H., DIMITRAKOPOULOU-STRAUSS, A., HEICHEL, T. O., LEHNER, B., BERND, L., EWERBECK, V., BURGER, C. & STRAUSS, L. G. 2001. Quantitative evaluation of skeletal tumours with dynamic FDG PET: SUV in comparison to Patlak analysis. *Eur J Nucl Med,* 28**,** 704-10.

WU, H. M., HUANG, S. C., CHOI, Y., HOH, C. K. & HAWKINS, R. A. 1995. A modeling method to improve quantitation of fluorodeoxyglucose uptake in heterogeneous tumor tissue. *J Nucl Med,* 36**,** 297-306.

YAREMKO, B., RIAUKA, T., ROBINSON, D., MURRAY, B., MCEWAN, A. & ROA, W. 2005. Threshold modification for tumour imaging in non-small-cell lung cancer using positron emission tomography. *Nucl Med Commun,* 26**,** 433-40.

YOUNG, H., BAUM, R., CREMERIUS, U., HERHOLZ, K., HOEKSTRA, O., LAMMERTSMA, A. A., PRUIM, J. & PRICE, P. 1999. Measurement of clinical and subclinical tumour response using [18F]-fluorodeoxyglucose and positron emission tomography: review and 1999 EORTC recommendations. European Organization for Research and Treatment of Cancer (EORTC) PET Study Group. *Eur J Cancer,* 35**,** 1773-82.

ZAFRA, M., AYALA, F., GONZALEZ-BILLALABEITIA, E., VICENTE, E., GONZALEZ-CABEZAS, P., GARCIA, T., MACIAS, J. A. & VICENTE, V. 2008. Impact of whole-body F-18-FDG PET on diagnostic and therapeutic management of Medical Oncology patients. *European Journal of Cancer,* 44**,** 1678-1683.

ZANZONICO, P. 2008. Routine quality control of clinical nuclear medicine instrumentation: a brief review. *J Nucl Med,* 49**,** 1114-31.

ZASADNY, K. R., KISON, P. V., FRANCIS, I. R. & WAHL, R. L. 1998. FDG-PET Determination of Metabolically Active Tumor Volume and Comparison with CT. *Clin Positron Imaging,* 1**,** 123-129.

ZASADNY, K. R. & WAHL, R. L. 1993. Standardized uptake values of normal tissues at PET with 2-[fluorine-18]-fluoro-2-deoxy-D-glucose: variations with body weight and a method for correction. *Radiology,* 189**,** 847-50.

ZHANG, H., WROBLEWSKI, K., APPELBAUM, D. & PU, Y. 2012. Independent prognostic value of whole-body metabolic tumor burden from FDG-PET in non-small cell lung cancer. *Int J Comput Assist Radiol Surg*.

ZIEGLER, S. I. 2005. Positron Emission Tomography: Principles, Technology, and Recent Developments. *Nuclear Physics A,* 752**,** 679c–687c.

ZIEPERT, M., HASENCLEVER, D., KUHNT, E., GLASS, B., SCHMITZ, N., PFREUNDSCHUH, M. & LOEFFLER, M. 2010. Standard International prognostic index remains a valid predictor of outcome for patients with aggressive CD20+ B-cell lymphoma in the rituximab era. *J Clin Oncol,* 28**,** 2373-80.

ZINZANI, P. L., GANDOLFI, L., BROCCOLI, A., ARGNANI, L., FANTI, S., PELLEGRINI, C., STEFONI, V., DERENZINI, E., QUIRINI, F. & BACCARANI, M. 2011. Midtreatment 18F-fluorodeoxyglucose positron-emission tomography in aggressive non-Hodgkin lymphoma. *Cancer,* 117**,** 1010-8.

ZITOVA, B. & FLUSSER, J. 2003. Image registration methods: a survey. *Image and Vision Computing,* 21**,** 977–1000.

ZSOTER, N., BANDI, P., GARAI, I. & PAPP, L. 2011. Hextuple registration of interim and follow-up PET-CT images for the accurate tracking of patient recovery after therapy. *Conf Proc IEEE Eng Med Biol Soc,* 2011**,** 2630-3.